AD_____

Award Number: DAMD17-03-1-0314

TITLE: Computer-Aided Detection of Mammographic Masses in Dense Breast Images

PRINCIPAL INVESTIGATOR: Lisa M. Kinnard

CONTRACTING ORGANIZATION: Howard University
Washington, DC 20059

REPORT DATE: June 2004

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 2004	Annual Summary (1 Jun 03-31 May 04)

4. TITLE AND SUBTITLE
Computer-Aided Detection of Mammographic Masses in Dense Breast Images

5. FUNDING NUMBERS
DAMD17-03-1-0314

6. AUTHOR(S)
Lisa M. Kinnard

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Howard University
Washington, DC 20059

E-Mail: l_kinnard@yahoo.com

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

This document describes the research tasks and educational activities in which the PI has been engaged during the first phase of this post-doctoral work. The PI has selected 342 dense breast masses (175 cancer cases and 167 benign cases) for testing using an automated segmentation algorithm which combines region growing with cost function analysis. This method has been validated on 198 cases using overlap, accuracy, sensitivity and specificity statistics by one expert radiologist and has been validated on 46 cases by a second expert radiologist. All statistics have values that range from 0 to 1. The mean overlap and accuracy values for cancer cases are approximately 0.46 and 0.78, respectively. The mean overlap and accuracy values for benign cases are approximately 0.52 and 0.87, respectively. The PI has attended cancer imaging workshops, attended two scientific meetings, made two oral presentations, team taught an Imaging Technologies course, and published two manuscripts during this time.

14. SUBJECT TERMS
Image Segmentation, Validation Methods, CAD

15. NUMBER OF PAGES
52

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

I. INTRODUCTION

Breast mass segmentation is arguably one of the most difficult tasks in the development of Computer-Aided Diagnostic (CAD_x) systems. The main objective of this research is to develop an image segmentation method for mammograms that contain dense tissue as well as for mammograms that contain dense/fatty tissue, while its second objective is to incorporate the segmentation method into a CAD_x system. Specifically, we intend to do the following: (1) To develop an automatic image segmentation scheme to separate clinically occult breast masses from surrounding tissue (2) To evaluate the method by comparing the ROIs with mammographers' drawings and (3) To separate masses from glandular tissues using the Multiple Circular Path Convolution Neural Network (MCPCNN) classifier.

II. BODY

During the past 12 months the PI has tested an automatic image segmentation algorithm on a set of dense breast mass cases. This section of the annual summary provides a detailed description of the experiment and is divided into the following sections: (A) Segmentation Method – an overview of the automated image segmentation method (please see Appendix for detailed description of method) (B) Database and Experiments – description of masses used and experiments performed (C) Results – statistical and graphical results of the experiment (D) Discussion of Results and (E) Future Work.

A. Segmentation Method

The segmentation method used in this study evaluates the steepest changes within a probabilistic cost function in an effort to determine the computer segmented contour which is most closely correlated with expert radiologist manual traces. It segments breast masses by combining region growing with the analysis of a probability-based function [1]. Once a set of contours is grown using region growing the probability density functions inside and outside the contours are found. A function, which is the logarithm of these probability density functions, is then constructed. The function is then searched for possible steep change locations, i.e., sharp changes in the logarithm values, and the intensities corresponding to those locations are likely to produce contours which are highly correlated with expert traces. A detailed description of the method is provided in the manuscripts located in the appendix of this document [2, 3].

B. Database and Experiments

The PI has selected 342 cases from the University of South Florida's Digital Database for Screening Mammography (DDSM) [4], where 175 of these cases are cancerous masses and 167 of the cases are benign masses. The densities of all cases from the DDSM have been rated according to the American College of Radiology's (ACR) density scale, which ranges from 1-4. A breast containing a great deal of fatty tissue would receive a rating of 1 and a breast containing a great deal of dense tissue would receive a rating of 4. The current database contains 242 cases with a density rating of 3 and 100 cases with a density rating of 4. In the current experiment the cost likelihood function threshold values (TV_1 and TV_2) were set to 1800 and 1300, respectively. Approximately 300 of the cases were manually traced by one expert radiologist (Expert 1), while 46 of the cases were manually traced by a second expert radiologist (Expert 2). One hundred ninety eight of the cases have been validated by one expert radiologist, while 45 of the cases have been validated by the second expert radiologist. The validation statistics are overlap, accuracy, sensitivity, and specificity as described in the manuscripts [2, 3].

C. Results

1. Statistical Results

As described in the manuscripts [2, 3], the cost likelihood function was used to narrow a large set of contour choices to three contours which are highly correlated with the expert radiologist traces. Figures 1-4 are plots of the percentages of cases regarding overlap and accuracy statistics that fall between specific data ranges for all three contour choices, namely, groups 1, 2, and 3. For example, the light gray bar corresponding to the label "0-0.09" shows the percentage of cases for the group 1 trace that fall between 0 and 0.09. All masses for the plots shown used Expert 1's traces for validation.

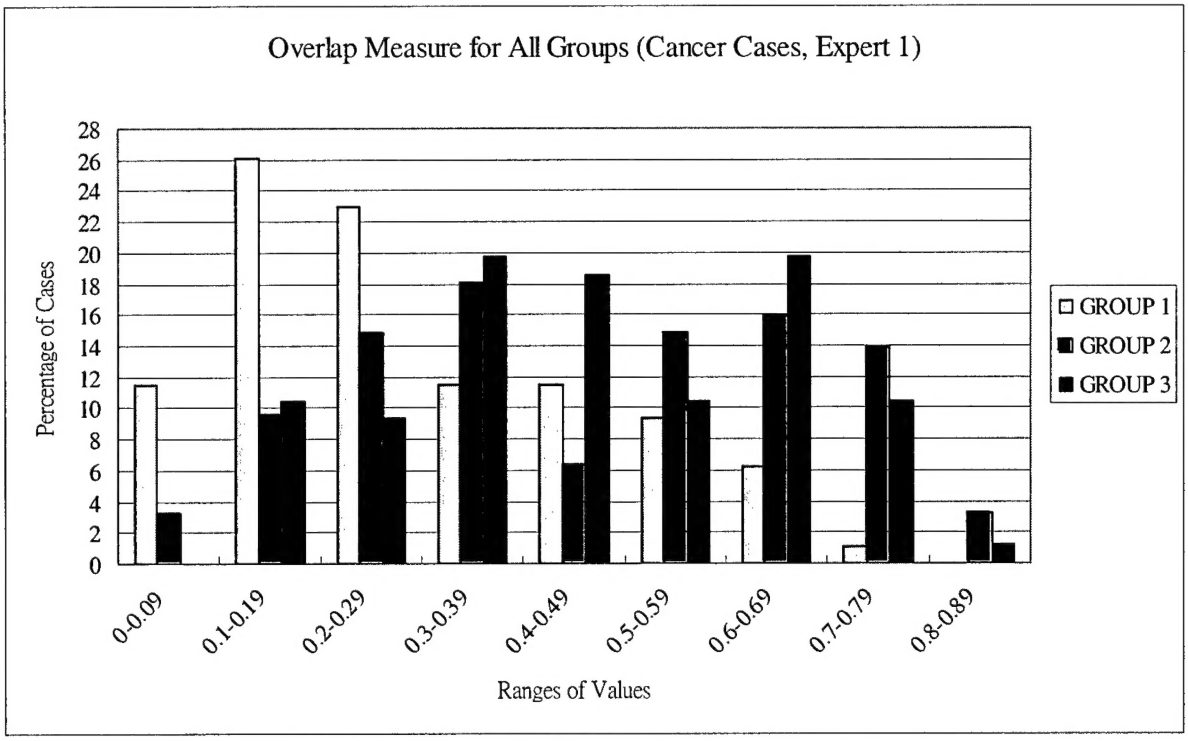


Figure 1 – Percentage of Cases for Ranges of Overlap Values (Cancer Cases)

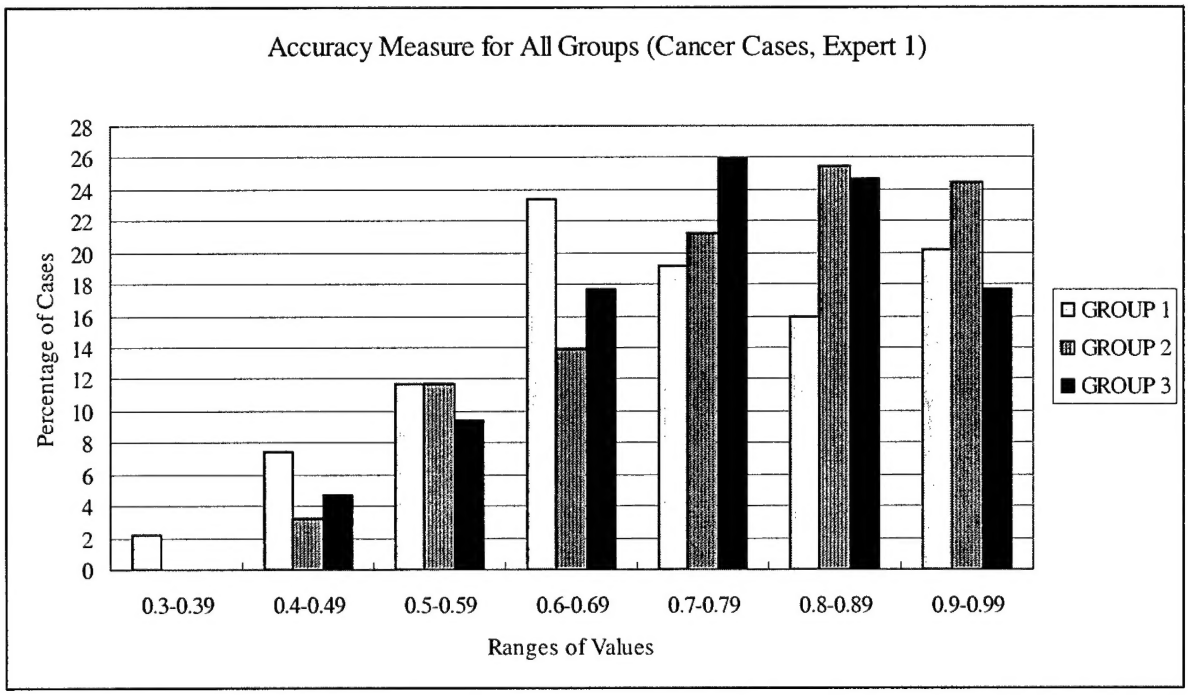


Figure 2 – Percentage of Cases for Ranges of Accuracy Values (Cancer Cases)

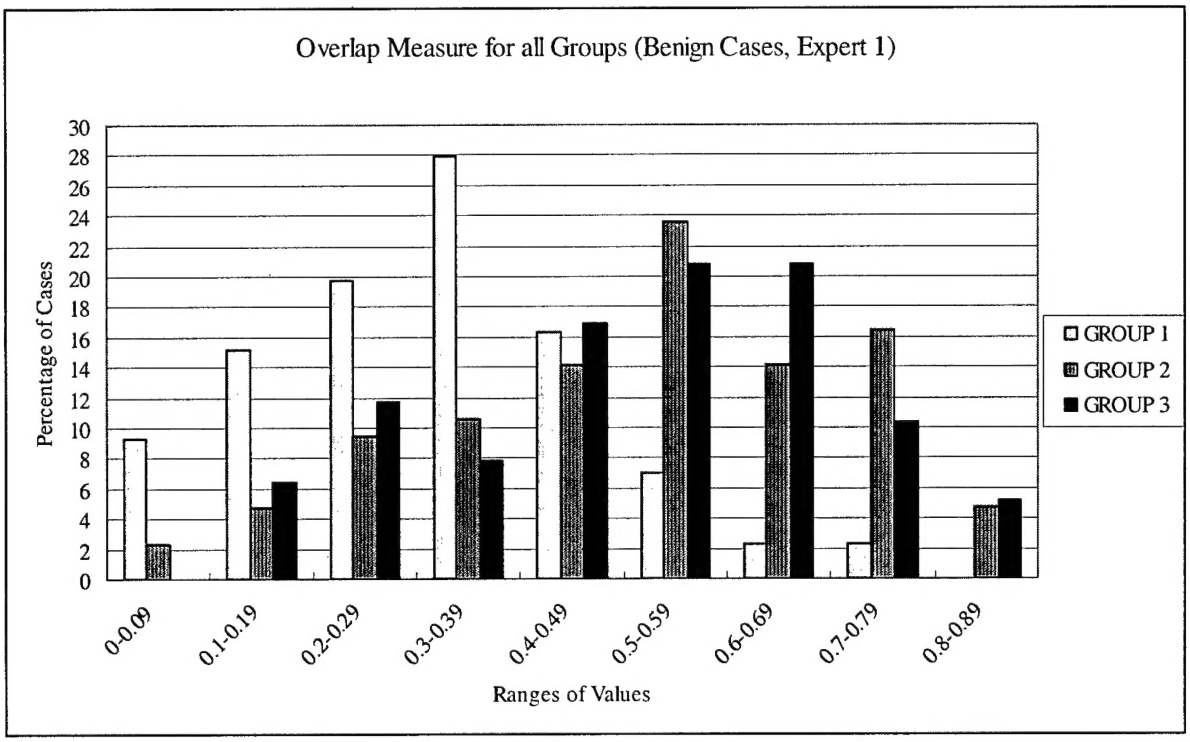


Figure 3 – Percentage of Cases for Ranges of Overlap Values (Benign Cases)

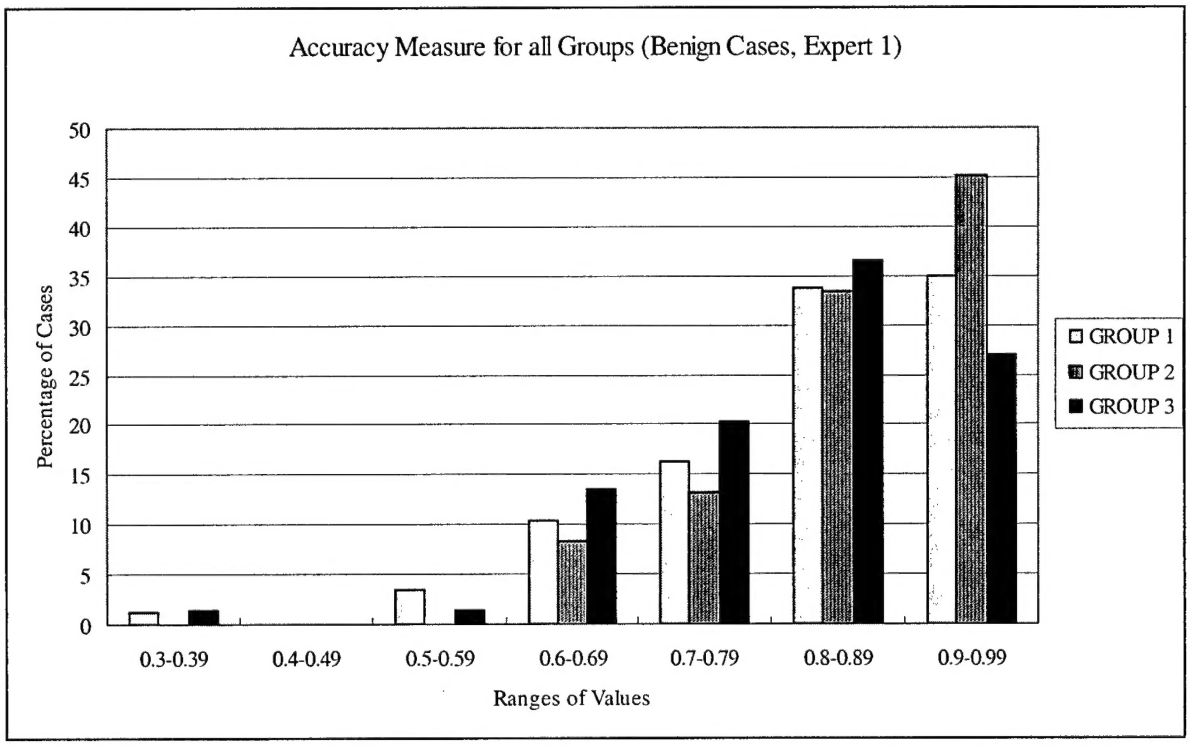


Figure 4 – Percentage of Cases for Ranges of Accuracy Values (Benign Cases)

Table 1 contains mean values of overlap and accuracy statistics for all contour groups for both cancerous and benign masses.

Table 1- Mean Value of Overlap and Accuracy Statistics for Groups 1, 2, and 3 (Expert 1)

	Cancer Cases		**Benign Cases**	
	Overlap	**Accuracy**	**Overlap**	**Accuracy**
Group 1	0.29	0.73	0.32	0.82
Group 2	0.46	0.78	0.52	0.87
Group 3	0.46	0.76	0.51	0.82

2. Visual Results

Figures 5-10 show results for cases in which there were strong, average, and low correlations between the computer-segmented results and Expert 1's manual traces for both cancer and benign cases. In cases for which Expert 2 has not yet traced the mass, the figure will indicate "no data". Figures 11 and 12 show results for cases where there is a great deal of disagreement between the Expert 1 and Expert 2 traces.

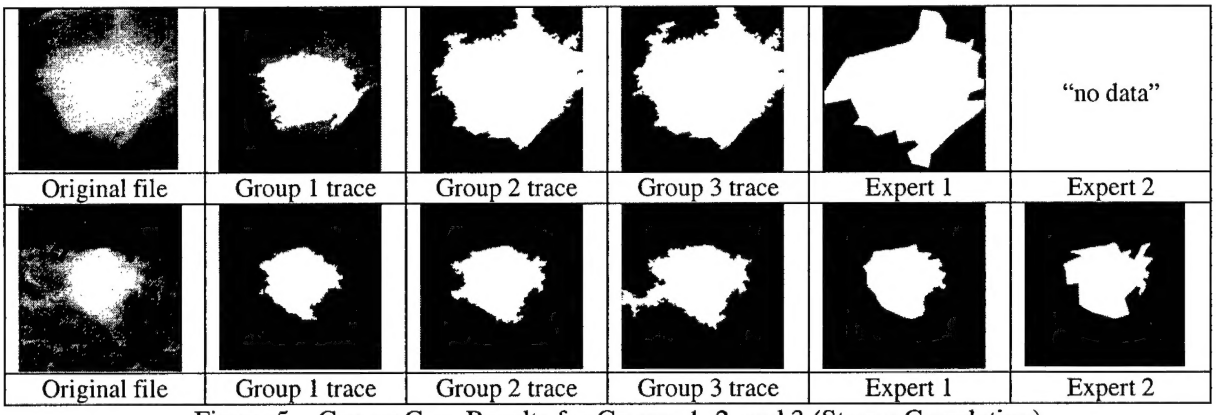


Figure 5 – Cancer Case Results for Groups 1, 2, and 3 (Strong Correlation)

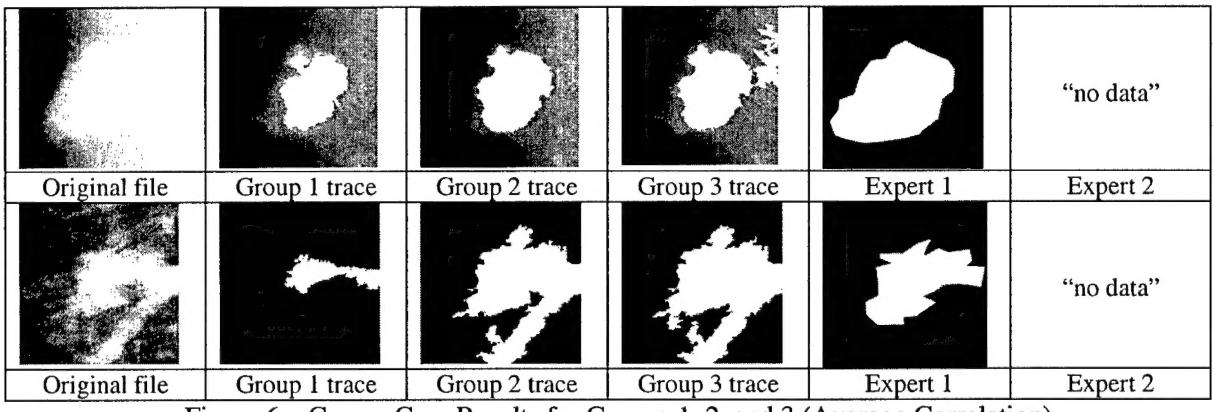


Figure 6 – Cancer Case Results for Groups 1, 2, and 3 (Average Correlation)

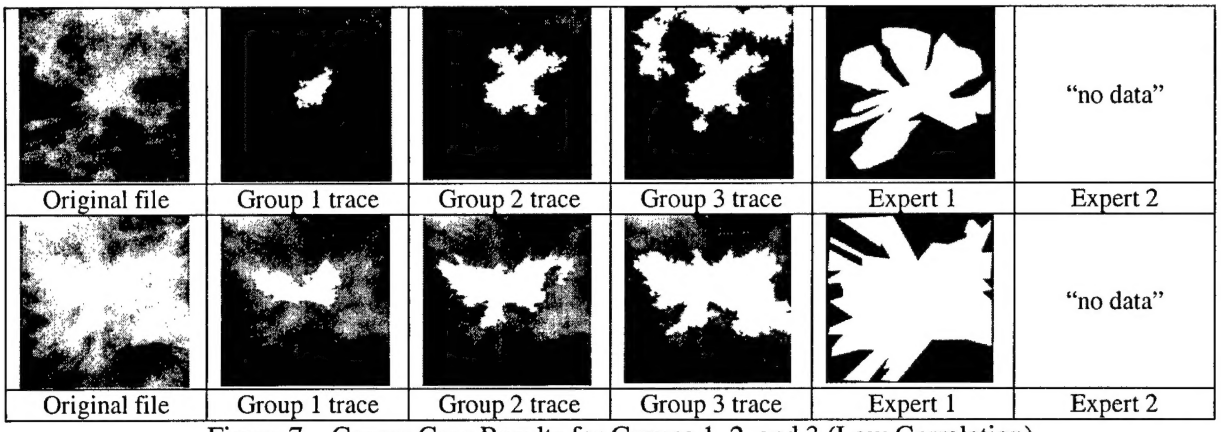


Figure 7 – Cancer Case Results for Groups 1, 2, and 3 (Low Correlation)

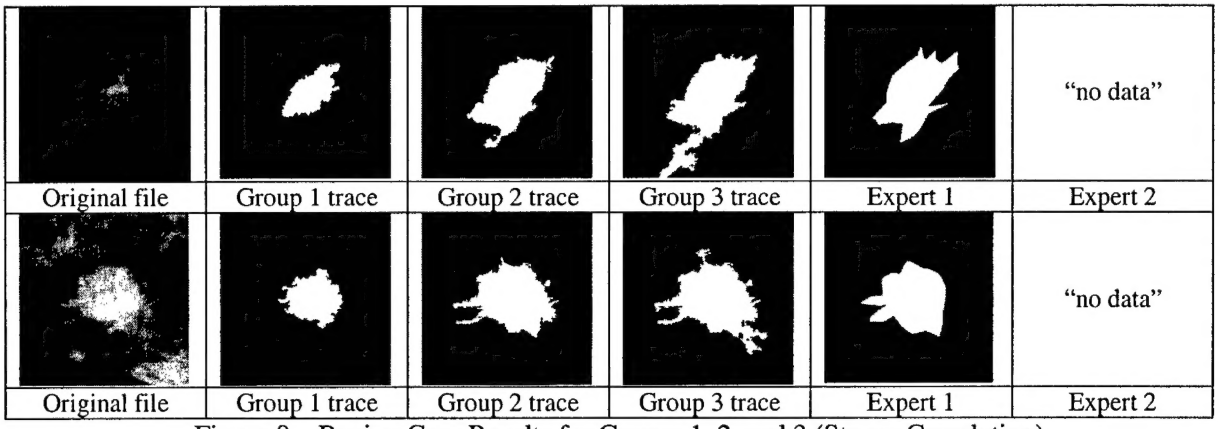


Figure 8 – Benign Case Results for Groups 1, 2, and 3 (Strong Correlation)

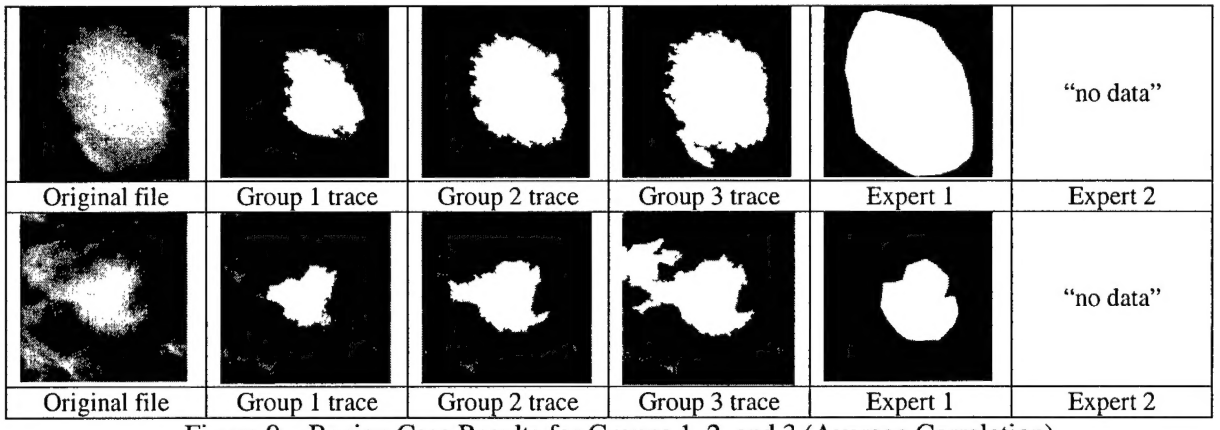


Figure 9 – Benign Case Results for Groups 1, 2, and 3 (Average Correlation)

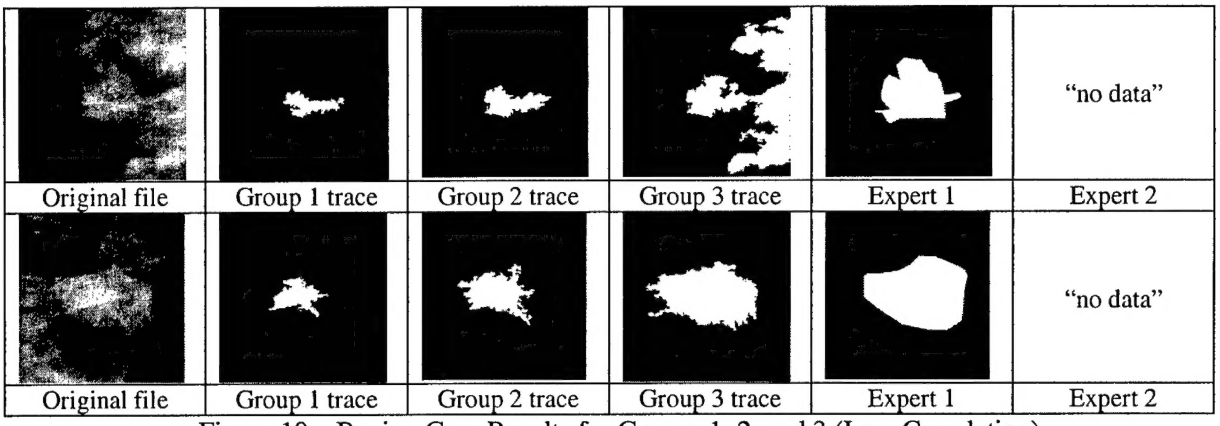


Figure 10 – Benign Case Results for Groups 1, 2, and 3 (Low Correlation)

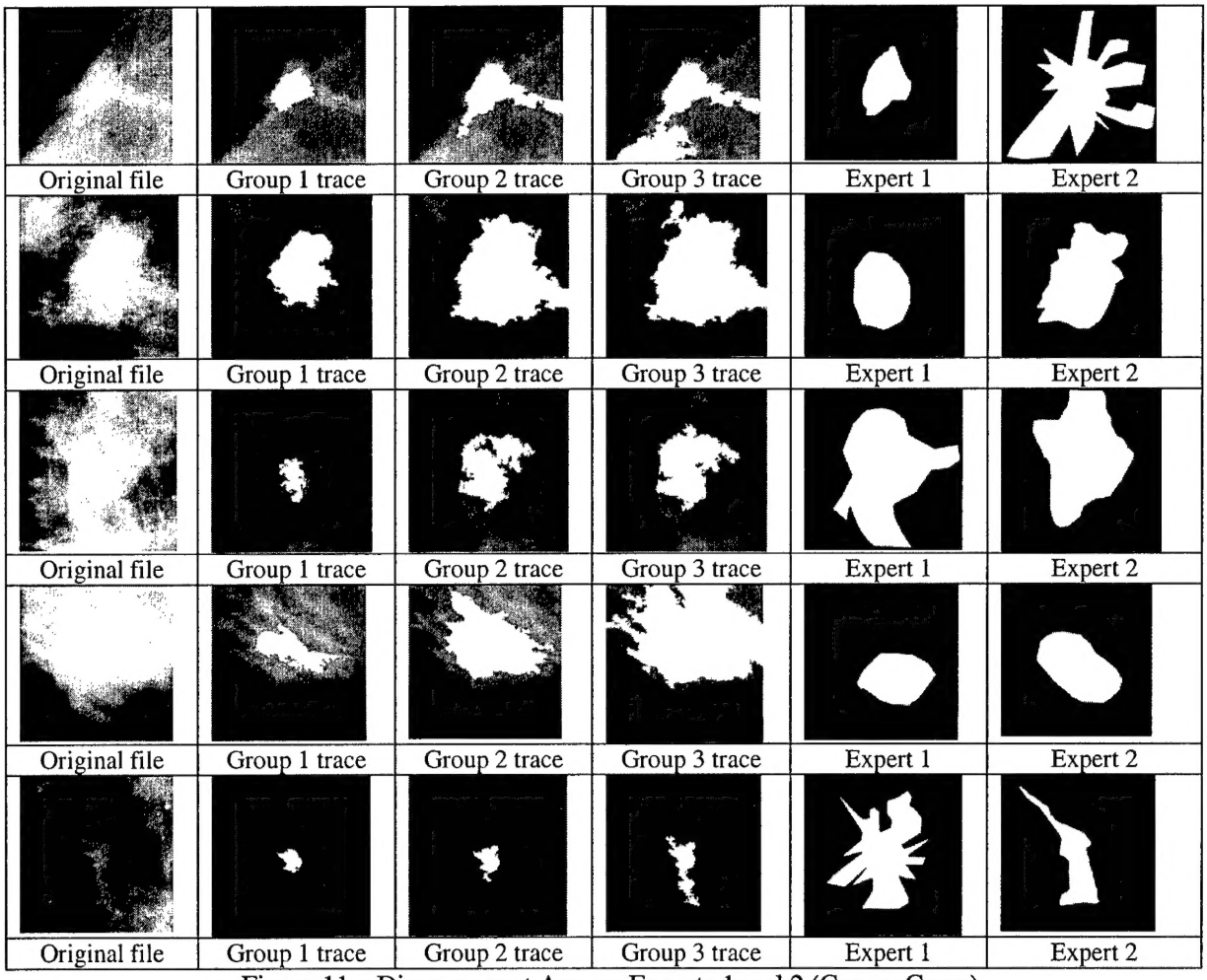


Figure 11 – Disagreement Among Experts 1 and 2 (Cancer Cases)

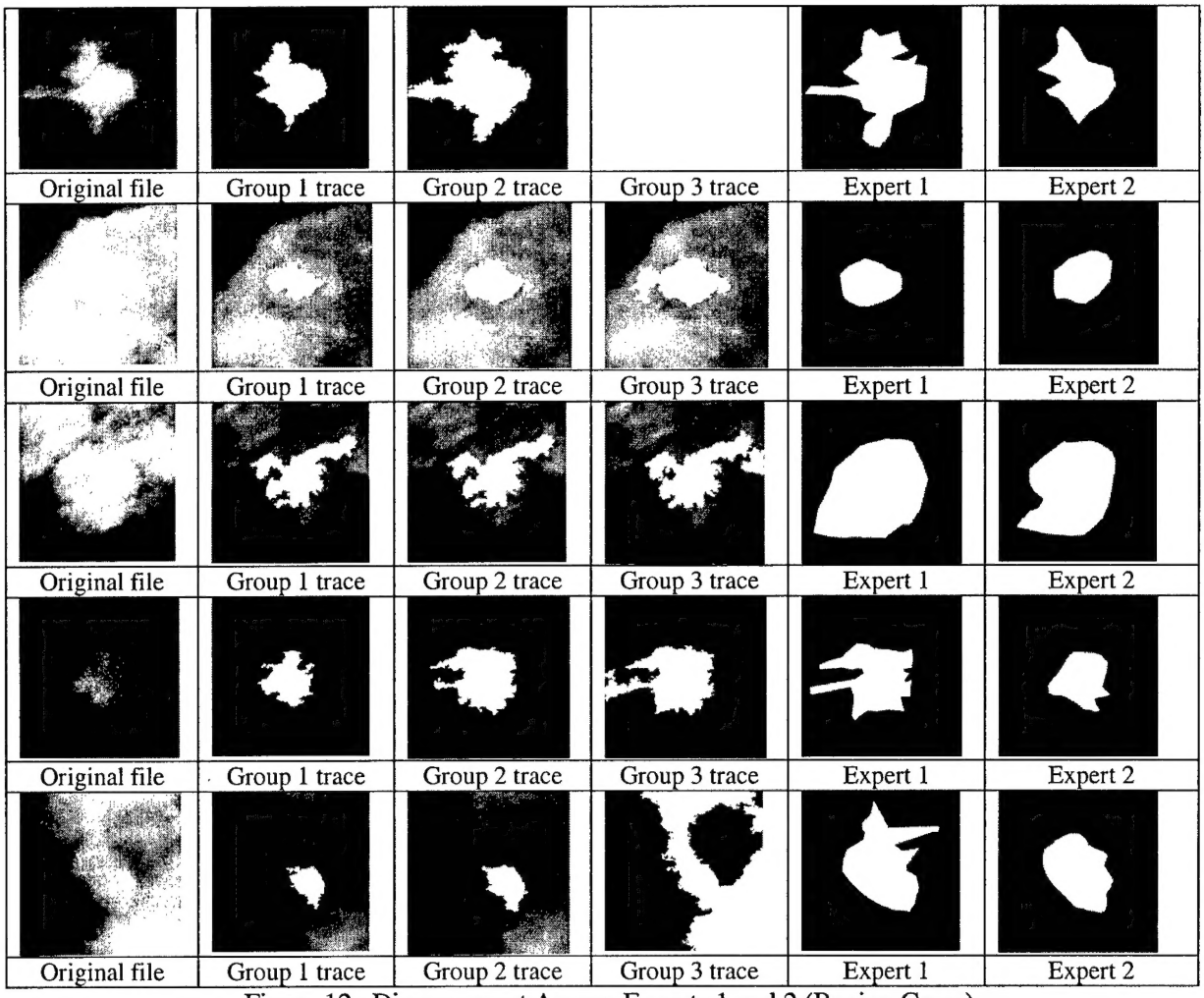


Figure 12 –Disagreement Among Experts 1 and 2 (Benign Cases)

D. Discussion of Results

Previous experiments showed that in most cases, the contour produced by the intensity corresponding to the first steep change location in the cost function best matched the expert radiologist traces regarding the overlap and accuracy statistics. This finding was verified *via* Analysis of Variance (ANOVA) testing, where p-values ranged from $1.03 \times 10^{-2} - 7.51 \times 10^{-17}$. However, the mean values listed in table 1 of this experiment revealed that the group 2 and group 3 traces performed equally well regarding the overlap statistic. The accuracy results were slightly higher for group 2 in comparison to group 3. Figures 1-4 reveal that there was a larger percentage of cases for which the group 2 and group 3 traces achieved higher overlap and accuracy values (approximately 0.6 and above), in comparison to the group 1 traces which achieved lower overlap and accuracy values (approximately 0.5 and below). This leads us to believe that the group 2 and group 3 traces generally match expert manual traces in the best way. This observation is consistent with the visual results shown in Figures 5-10. We have also obtained higher values for overlap and accuracy values in previous experiments because the boundaries of the masses in these data sets were clearer than those for the dense mass cases.

We also recognize that for cases in which the intensity values outside the perceived boundaries of the mass are greater than or equal to values inside the mass, the region growing technique produces contours that have grown into those areas. We refer to this phenomenon as flooding because the contour "floods" into fibroglandular tissue that is not actual mass tissue. Furthermore, areas inside the mass that have low

intensity values are excluded from contour growth so some mass tissue is missed during the region growing phase. These are limitations of the current segmentation algorithm and we plan to address these issues during the next phase of this research work.
Figures 11 and 12 indicate that large differences in expert interpretations of mass boundaries can exist, therefore we are motivated us to raise questions around the consistency and reliability of expert traces. In some cases there is strong agreement between the computer trace and Expert 1's trace while in other cases there is strong agreement between the computer trace and Expert 2's trace, so it is important to analyze expert opinions prior to further optimizing the segmentation method, regarding image filtering.

E. Future Work

The statement of work states that the PI would filter the images following the experiment performed during the past several months, however, the question of expert reliability has come into question and should be addressed prior to the optimization of the algorithm. During the next phase of this research work, the PI will compare the computer-segmented results to each observer as well has compare the experts' traces to each other in efforts to answer the questions around expert reliability and consistency.

III. KEY RESEARCH ACCOMPLISHMENTS

- Selected nearly 350 mass cases with density ratings of 3 and 4 from the University of South Florida's Digital Database for Screening Mammography
- Collected approximately 100 cases from the Georgetown University Medical Center image database
- 300 masses have been delineated by one expert radiologist
- 46 masses have been delineated by a second expert radiologist (this radiologist has agreed to delineate the remaining cases)
- Validated 198 masses regarding overlap, accuracy, sensitivity, and specificity statistics
- Reviewed literature regarding reliability and consistency among expert observers [5-10]

IV. REPORTABLE OUTCOMES

Manuscripts:

1. L. Kinnard, S.-C. B. Lo, E. Makariou, T. Osicka, P. Wang, M.T. Freedman, M. Chouikha, "Likelihood Function Analysis for Segmentation of Mammographic Masses for Various Margin Groups", Proceedings of the IEEE Symposium on Biomedical Imaging, April 2004.
2. L. Kinnard, S.-C. B. Lo, E. Makariou, T. Osicka, P. Wang, M.T. Freedman, M. Chouikha, "Steepest changes of a probability-based cost function for delineation of mammographic masses: A validation study", Medical Physics (manuscript submitted 12/03, revised manuscript submitted 4/04, revised **manuscript accepted 6/04**)
3. L. Kinnard, Ph.D. thesis, "Segmentation of Mass Bodies and Their Extended Borders on Mammograms Using Maximum-Likelihood Analysis", June 2003.

Oral Presentations:

1. L. Kinnard, S.-C. B. Lo, E. Makariou, T. Osicka, P. Wang, M.T. Freedman, M. Chouikha, "Likelihood Function Analysis for Segmentation of Mammographic Masses for Various Margin Groups", Proceedings of the IEEE Symposium on Biomedical Imaging, April 2004.
2. "Breast Cancer Research: Computer-Aided Diagnosis and Image Segmentation", Howard University Cancer Center, December 2003.

Technical Development Activities:

- Attended two cancer imaging workshops sponsored by the Washington Academy of Biomedical Engineering (WABME):
    1. 9/29/03: "Individualized Treatment Using Pharmaco-Genomics & Functional Imaging" (George Washington University)

2. 11/12/03: "Cancer Imaging for the Operating Room of 2020" (Georgetown University)

- Attended weekly cancer workshops conducted by the Howard University Cancer Center
- Attended SPIE Medical Imaging Meeting (February, 2004, San Diego, CA)
- Taught "Computer-Aided Diagnosis" portion of "Introduction to Imaging Technologies" course, The Catholic University of America (course number ENGR552)

V. CONCLUSIONS

Overall the segmentation method has produced overlap and accuracy values that reflect the difficulties of the data set. These values are generally lower than those of previous experiments because the boundaries of dense tissue masses are exceedingly difficult to locate. The flooding phenomenon is also responsible for these low values because in some cases the intensity values of the masses are very close to those of surrounding fibroglandular tissue. The difficulty of locating mass boundaries is also reflected in some cases where there are significant differences between the expert traces. The computer-segmented results are sometimes more closely correlated with one expert in some cases while the results are more closely correlated with the second expert in other cases. During the next phase of this research work, the PI will compare the computer-segmented results to each observer as well has compare the experts' traces to each other in efforts to answer the questions around expert reliability and consistency. Once these questions have been answered we believe that it will possible to decide how the images need to be filtered or if they need to be filtered at all.

VI. REFERENCES

1. M.A. Kupinski, M.L. Giger, "Automated Seeded Lesion Segmentation on Digital Mammograms", IEEE Trans. on Med. Imag., **17**, no. 4, 510-517 (1998).
2. L. Kinnard, S.-C. B. Lo, E. Makariou, T. Osicka, P. Wang, M.T. Freedman, M. Chouikha, "Likelihood Function Analysis for Segmentation of Mammographic Masses for Various Margin Groups", Proceedings of the IEEE Symposium on Biomedical Imaging, 113-116 (2004).
3. L. Kinnard, S.-C. B. Lo, E. Makariou, T. Osicka, P. Wang, M.T. Freedman, M. Chouikha, "Steepest changes of a probability-based cost function for delineation of mammographic masses: A validation study", Medical Physics (manuscript submitted 12/03, revised manuscript submitted 4/04, revised **manuscript accepted 6/04**)
4. M. Heath, K.W. Bowyer, D. Kopans et al., "Current status of the digital database for screening mammography", Digital Mammography, Kluwer Academic Publishers, 457-460 (1998).
5. V. Chalana, Y. Kim, "A Methodology for Evaluation of Boundary Detection Algorithms on Medical Images", IEEE Trans. on Med. Imag., **16**, no. 5, 642-652 (1997).
6. P. Besl, N.D. McKay, "A Method for Registration of 3-D Shapes", IEEE Trans. on Med. Imag., **14**, no. 2, 239-255 (1992).
7. I.M. Anderson, J.C. Bezdek, "Curvature and Tangential Deflection of Discrete Arcs: A Theory Based on the Commutator of Scatter Matrix Pairs and Its Application to Vertex Detection in Planar Shape Data", IEEE Trans. on Pattern Analysis and Machine Intelligence, **PAMI-6**, no. 1, 27-40 (1984).
8. J.M. Bland, D.G. Altman, "Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement", The Lancet, **1**, 307-310 (1986).
9. F.J. Rohlf, D. Slice, "Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks", Systematic Zoology, **39**, no. 1, 40-59 (1990).
10. F. Bookstein, "Landmark methods for forms without landmarks: morphometrics of group differences in outline shape", Medical Image Analysis, **1**, no. 3, 225-243 (1996/7).

LIKELIHOOD FUNCTION ANALYSIS FOR SEGMENTATION OF MAMMOGRAPHIC MASSES FOR VARIOUS MARGIN GROUPS

Lisa Kinnard[a,b,c], Shih-Chung B. Lo[a], Erini Makariou[a], Teresa Osicka[a,d], Paul Wang[c], Matthew T. Freedman[a], Mohamed Chouikha[b]

[a]ISIS Center, Dept. of Radiology, Georgetown University Medical Center, Washington, D.C., USA
[b]Department of Electrical and Computer Engineering, Howard University, Washington, D.C., USA
[c]Biomedical NMR Laboratory, Department of Radiology, Howard University, Washington, D.C., USA
[d]Department of Electrical Engineering and Computer Science, The Catholic University of America, Washington DC, USA

ABSTRACT

The purpose of this work was to develop an automatic boundary detection method for mammographic masses and to observe the method's performance on different four of the five margin groups as defined by the ACR, namely, spiculated, ill-defined, circumscribed, and obscured. The segmentation method utilized a maximum likelihood steep change analysis technique that is capable of delineating ill-defined borders of the masses. Previous investigators have shown that the maximum likelihood function can be utilized to determine the border of the mass body. The method was tested on 122 digitized mammograms selected from the University of South Florida's Digital Database for Screening Mammography (DDSM). The segmentation results were validated using overlap and accuracy statistics, where the gold standards were manual traces provided by two expert radiologists. We have concluded that the intensity threshold that produces the best contour corresponds to a particular steep change location within the likelihood function.

1. INTRODUCTION

In a CAD_x system, segmentation is arguably one of the most important aspects – particularly for masses – because strong diagnostic predictors for masses are shape and margin type [2,9]. The margin of a mass is defined as the interface between the mass and surrounding tissue [2]. Furthermore, breast masses can have unclear borders and are sometimes obscured by glandular tissue in mammograms. A spiculated mass consists of a central mass body surrounded by fibrous projections, hence the resulting stellate shape. For the aforementioned reasons, proper segmentation - to include the body and periphery - is extremely important and is essential for the computer to analyze, and in turn, determine the malignancy of the mass in mammographic CAD_x systems.

Over the years researchers have used many methods to segment masses in mammograms. Petrick [7] et al. developed the Density Weighted Contrast Enhancement (DWCE) method, in which series of filters are applied to the image in an attempt to extract masses. Comer et al. [1] segmented digitized mammograms into homogeneous texture regions by assigning each pixel to one of a set of classes such that the number incorrectly classified pixels was minimized via Maximum Likelihood (ML) analysis. Li [5] developed a method that employs k-means classification to classify pixels as belonging to the region of interest (ROI) or background.

Kupinski and Giger developed a method [4], which uses ML analysis to determine final segmentation. In their method, the likelihood function is formed from likelihood values determined by a set of image contours produced by the region growing method. This method is a highly effective one that was also implemented by Te Brake and Karssemeijer in their comparison between the discrete dynamic contour model and the likelihood method [9]. For this reason we chose to investigate its use as a possible starting point from which a second method could be developed. Consequently in our implementation of this work we discovered an important result, i.e., the maximum likelihood steep change. It appears that in many cases this method produces contour choices that encapsulate important borders such as mass spiculations and ill-defined borders.

2. METHODS

2.1 Initial Contours

As an initial segmentation step, we followed the overall region similarity concept to aggregate the area of interest [1, 4]. Used alone, a sequence of contours representing the mass is generated; however, the computer is not able to choose the contour that is most closely correlated with the experts' delineations. Furthermore, we have devised an ML function steep change analysis method that chooses the best contour that delineates the mass body as well as its extended borders, i.e., extensions into spiculations and areas in which the borders are ill-defined or obscured. This method is an extension of the method developed by Kupinski and Giger [4] that uses ML function analysis to select the contour which best represents the mass, as compared to expert radiologist traces. We have determined that this technique can select the contour that accurately represents the mass body contour for a given set of parameters; however, further analysis of the likelihood function revealed that the computer could

0-7803-8388-5/04/$20.00

choose a set of three segmentation contour choices from the entire set of contour choices, and then make a final decision from these three choices.

The algorithm can be summarized in several steps. Initially, we use an intensity based thresholding scheme to generate a sequence of grown contours (S_i), where gray value is the similarity criterion. The image is also multiplied by a 2D trapezoidal membership function (2D shadow), whose upper base measures 40 pixels and lower base measures 250 pixels (1 pixel = 50 microns). The image to which the shadow has been applied is henceforth referred to as the "fuzzy" image. The original image and its fuzzy version were used to compute the likelihood of the mass's boundaries. The computation method is comprised of two components for a given boundary: (1) formulation of the composite probability and (2) evaluation of likelihood.

In addition, we chose to aggregate contours using the original image. This accounts for the major difference from that implemented by the previous investigators. Since smoother contours were not used, the likelihood function showed greater variations. In many situations, the greatest variations occurred when there was a sudden increase of the likelihood, and this was strongly correlated with the end of the mass border growth. This phenomenon would be suppressed if the fuzzy image was used to generate the contours. The fuzzy image was used mainly to construct the likelihood function.

2.2 Composite Probability Formation

For a contour (S_i), the composite probability (C_i) is calculated:

$$C_i|S_i = p(f_i(x,y)|S_i) \times p(m_i(x,y)|S_i) \qquad (1)$$

The quantity $f_i(x,y)$ is the area to which the 2D shadow has been multiplied, $p(f_i(x,y)|S_i)$ is the probability density function of the pixels inside S_i where 'i' is the region growing step associated with a given intensity threshold. The quantity $m_i(x,y)$ is the area outside S_i (non-fuzzy), and $p(m_i(x,y)|S_i)$ is the probability density function of the pixels outside S_i. Next we find the logarithm of the composite probability of the two regions, C_i:

$$Log(C_i|S_i) = \log(p(f_i(x,y)|S_i)) + \log(p(m_i(x,y)|S_i)) \qquad (2)$$

2.3 Evaluation of Likelihood Function

The likelihood that the contour represents the fibrous portion of the mass, i.e., mass body is determined by assessing the maximum likelihood function:

$$\arg\max(Log(C_i|S_i)), S_i, i = 1,...n \qquad (3)$$

Equation (3) intends to find the maximum value of the aforementioned likelihood values as a function of intensity threshold. It has been assessed (also by other investigators [4]) that the intensity value corresponding to this maximum likelihood value is the optimal intensity needed to delineate the mass body contour. However, in our implementation it was discovered that the intensity threshold corresponding to the maximum likelihood value confines the contour to the mass body. In our study many of these contours did not include the extended borders. We, therefore, hypothesize that the contour represents the mass's extended borders may well be determined by assessing the maximum changes of the likelihood function, i.e., locate the steepest likelihood value changes within the function:

$$\frac{d}{di}(Log(C_i|S_i)), S_i, i = 1,\ldots,n \qquad (4)$$

Based on this assumption, we have carefully analyzed the behavior of maximum likelihood function. The analysis reveals that we have successfully discovered that the most accurate mass delineation is usually obtained by using the intensity value corresponding to the first or second steep change locations within the likelihood function immediately following the maximum likelihood value on the likelihood function.

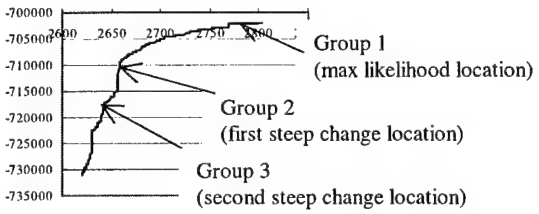


Figure 1: A likelihood function with steep change indicators

2.4 Steep change definition

The term "steep change" is rather subjective and can defined as a location between two or more points in the function where the likelihood values experience a significant change. In some cases the likelihood function increases at a slow rate. The algorithm design accounts for this issue by calculating the difference between likelihood values in steps over several values and comparing the results to two thresholds. The difference equation is given by:

$$h(t) = f(z - wt) - f(z - w(t+1)), \quad t = 0,\ldots,N \qquad (5)$$

where f is the likelihood function, z is the maximum intensity, w is the width of the interval over which the likelihood differences are calculated (e.g. – for w=7 differences are calculated every 7 points), and N is the total number of points in the searchable area divided by w. If the calculation in question yields a value greater than or equal to a given threshold, then the intensity corresponding to this location is considered to be a steep change location. The threshold algorithm occurs as follows:

If $(h(t)_{ML} \geq ML_{T1})$; t=0,...,m
Then choice 1 = intensity where that condition is satisfied
If $(h(t)_{ML} \geq ML_{T2})$; t=m,...,z
Then choice 2 = intensity where that condition is satisfied

where $h(t)_{ML}$ is the steep change value given by equation (5), ML_{T1} and ML_{T2} are pre-defined threshold values, m is the location in the function where the choice 1 condition is satisfied, and z is the location in the function where the choice 2 condition is satisfied. Once the condition is satisfied for the first threshold value (ML_{T1}) then its corresponding intensity value is used to produce the segmentation contour for the first steep change location. Once the condition is satisfied for ML_{T2} then its corresponding intensity value is used to produce the segmentation contour for the second steep change location.

2.5 Validation

The segmentation method was validated on the basis of overlap and accuracy [8,10]:

$$Overlap = \frac{N_{TP}}{N_{FN} + N_{TP} + N_{FP}} \qquad (6)$$

$$Accuracy = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}} \quad (7)$$

where N_{TP} is the true positive fraction, N_{TN} true negative fraction, N_{FP} is the false positive fraction, and N_{FN} is the false negative fraction. The gold standards used for the validation study were mass contours, which have been traced by expert radiologists.

Our experiments produced contours for the intensity values resulting from three locations within the likelihood functions: (1) The intensity for which a value within the likelihood function is maximum (group 1 contour) (2) The intensity for which the likelihood function experiences its first steep change (group 2 contour) and (3) The intensity for which the likelihood function experiences its second steep change (group 3 contour). We have observed that the intensity for which the likelihood function experiences its first steep change produces the contour trace that is most highly correlated with the gold standard traces, regarding overlap and accuracy.

3. EXPERIMENTS AND RESULTS

Here we describe the database used, describe the experiments, provide visual results obtained by the algorithm, as well as report the results obtained by the ANOVA test.

3.1 Database

For this study, a total of 122 masses were chosen from the University of South Florida's Digital Database for Screening Mammography (DDSM) [3]. The films were digitized at resolutions of 43.5 or 50 μm's using either the Howtek or Lumisys digitizers, respectively. The DDSM cases have been ranked by expert radiologists on a scale from 1 to 5, where 1 represents the most subtle masses and 5 represents the most obvious masses. The images were of varying subtlety ratings. The first set of expert traces was provided by an attending physician of the GUMC, and is hereafter referred to as the Expert A traces. The second set of expert traces was provided by the DDSM, and is hereafter referred to as the Expert B traces.

3.2 Experiments and Results

As mentioned previously, the term "steep change" is very subjective and therefore a set of thresholds needed to be set in an effort to define a particular location within the likelihood function as a "steep change location". For this study the following thresholds were experimentally chosen: ML_{T1}=1800, ML_{T2}=1300, where ML_{T1}= threshold for steep change location 1 for the likelihood function, and ML_{T2} = threshold for steep change location 2 for the likelihood function. We performed a number of experiments in an effort to prove that the intensity for which the likelihood function experiences the first steep change location produces the contour trace, which is most highly correlated with the gold standard traces regarding overlap and accuracy.

First we present segmentation results for two malignant cases followed segmentation results for two benign cases. Each figure contains an original image, traces for Experts A and B, and computer segmentation results for groups 1, 2, and 3. Second, we present data that plots the mean values for various margin groups for both overlap and accuracy measurements. The plots present data for the spiculated and ill-defined groups of malignant masses, and ill-defined and circumscribed groups of benign masses. Data was not presented for the other categories because there was not a sufficient amount of cases.

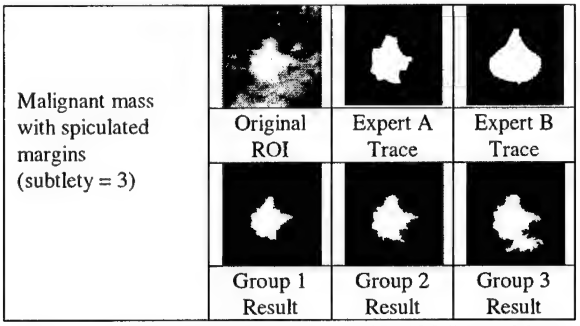


Figure 2: Segmentation Results: Spiculated Malignant Mass

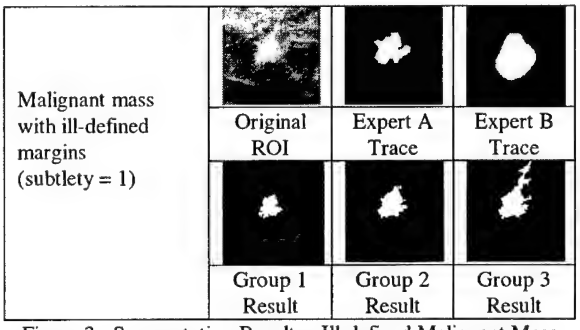


Figure 3: Segmentation Results: Ill-defined Malignant Mass

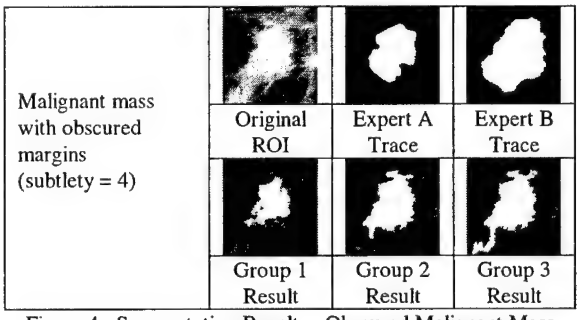


Figure 4: Segmentation Results: Obscured Malignant Mass

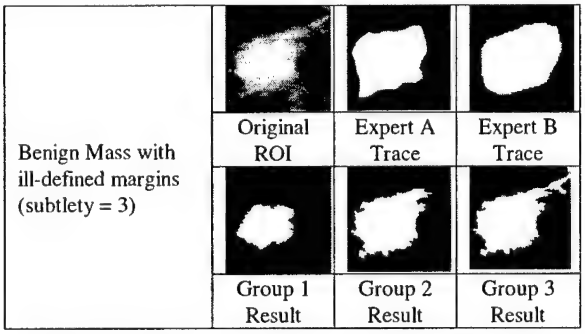


Figure 5: Segmentation Results: Ill-defined Benign Mass

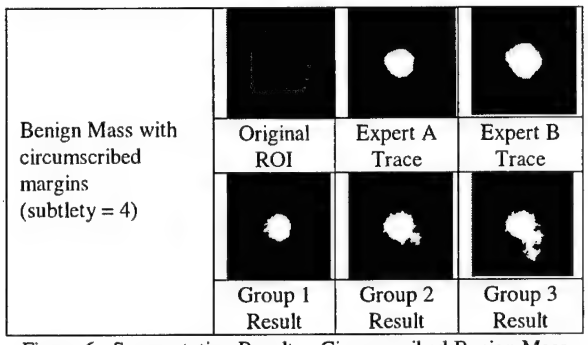


Figure 6: Segmentation Results: Circumscribed Benign Mass

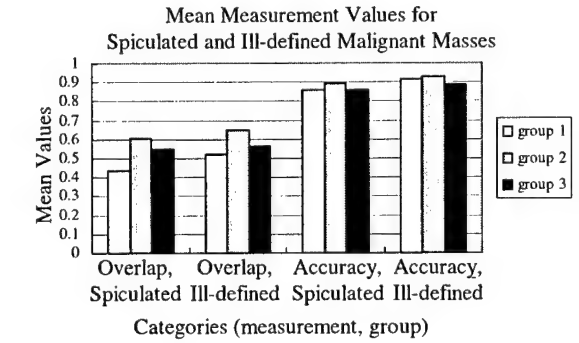


Figure 7: Mean Measurement Values (Malignant Masses)

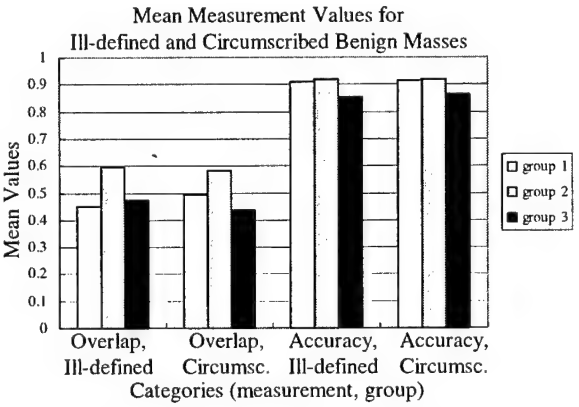


Figure 8: Mean Measurement Values (Benign Masses)

4. DISCUSSION AND CONCLUSION

The visual results (see Figures 2-6) reveal that the group 2 trace appears to delineate the masses better than the group 1 and group 3 contours in most cases. Visually, it appears that the method has performed equally well on all margin groups. This is an encouraging result because some of the more difficult masses to segment are typically those that are spiculated, obscured, and those that have ill-defined borders. The plots shown in Figures 7-8 confirm that the group 2 trace performs better than the other groups on the basis of overlap and accuracy for all margin groups, therefore supporting our visual observations.

In future work, a worthwhile study would be to test gather more data for all margin groups in an effort to see if the various groups require different parameter values to maximize the algorithm's robustness. Our ultimate goal is to optimize its performance for those masses falling in the ill-defined and obscured margin groups because segmentation of masses falling into those categories is exceedingly difficult.

5. REFERENCES

[1] M.L. Comer, E.J. Delp, "The EM/MPM algorithm for segmentation of textured images: Analysis and further experimental results", *Proceedings of the 1995 IEEE ICIP*, Lausanne, Switzerland, September 16-19, 1996.

[2] J.R. Harris, M.E. Lippman, M. Morrow, S. Hellman, "Diseases of the breast", Lippincott-Raven Publishers, Philadelphia, PA, pp. 80-81, 1996.

[3] M. Heath, K.W. Bowyer, D. Kopans et al., "Current status of the digital database for screening mammography", Digital Mammography, Kluwer Academic Publishers, pp. 457-460 , 1998.

[4] M.A. Kupinski, M.L. Giger, "Automated Seeded Lesion Segmentation on Digital Mammograms", *IEEE Trans. on Med. Imag.*, vol. 17, no. 4, pp. 510-517, 1998.

[5] L. Li, Y. Zheng, L. Zhang, R. Clark, "False-positive reduction in CAD mass detection using a competitive classification strategy", *Med. Phys.*, vol. 28, pp. 250-258, 2001.

[6] J.E. Martin, "Atlas of mammography: histologic and mammographic correlations (second edition)", Williams and Wilkins, Baltimore, MD, p. 87, 1988.

[7] N. Petrick, H-P Chan, B. Sahiner, D. Wei, "An Adaptive Density-Weighted Contrast Enhancement Filter for Mammographic Breast Mass Detection", *IEEE Trans. on Med. Imag.*, vol. 15, no. 1, pp. 59-67, 1996.

[8] J. Suckling, D.R. Dance, E. Moskovic, D.J. Lewis, S.G. Blacker, "Segmentation of mammograms using multiple linked self-organizing neural networks", *Med. Phys.*, vol. 22, pp. 145-152, 1995.

[9] G.M. te Brake, N. Karssemeijer, "Segmentation of suspicious densities in digital mammograms", *Med. Phys.*, vol. 28, no. 2, pp. 259-266, 2001.

[10] B. Van Ginneken, "Automatic segmentation of lung fields in chest radiographs", *Med. Phys.*, 27, pp. 2445-2455, 2000.

6. ACKNOWLEDGMENTS

This work was supported by US Army Grant numbers DAMD17-01-1-0267, DAMD 17-00-1-0291, DAAG55-98-1-0187, and DAMD 17-03-1-0314.

Steepest changes of a probability-based cost function for delineation of mammographic masses: A validation study

Lisa Kinnard[1,2,3], Shih-Chung B. Lo[1,2], Erini Makariou[1], Teresa Osicka[1,4], Paul Wang[3], Mohamed F. Chouikha[2], and Matthew T. Freedman[1]

[1]ISIS Center, Georgetown University Medical Center, Washington, D.C.
[2]Department of Electrical and Computer Engineering, Howard University, Washington, D.C.
[3]Biomedical NMR Laboratory, Department of Radiology, Howard University,Washington, D.C.
[4]Department of Electrical Engineering and Computer Science,
The Catholic University of America, Washington, D.C.

Submitted to J Medical Physics for consideration as a Technical Report

Please Send the Correspondence to:

Email: lo@isis.imac.georgetown.edu

Dr. Shih-Chung B. Lo
ISIS Center
Radiology Department
Georgetown University
Box 571479
Washington, D.C. 20057-1479

Abstract

The purpose of this work was to develop an automatic boundary detection method for mammographic masses and to rigorously test this method via statistical analysis. The segmentation method utilized a steepest change analysis technique for the determination of the mass boundaries based on a composed probability density cost function. Previous investigators have shown that this function can be utilized to determine the border of mass body. We have further analyzed this method and have discovered that the steepest changes in this function can produce mass delineations to include extended projections. The method was tested on 124 digitized mammograms selected from the University of South Florida's Digital Database for Screening Mammography (DDSM). The segmentation results were validated using overlap, accuracy, sensitivity, and specificity statistics, where the gold standards were manual traces provided by two expert radiologists. We have concluded that the best intensity threshold corresponds to a particular steepest change location within the composed probability density function. We also found that our results are more closely correlated with one expert than with the second expert. These findings were verified *via* Analysis of Variance (ANOVA) testing. The ANOVA tests obtained p-values ranging from $1.03 \times 10^{-2} - 7.51 \times 10^{-17}$ for the single observer studies, $2.03 \times 10^{-2} - 9.43 \times 10^{-4}$ for the two observer studies, and results were categorized using three significance levels, i.e., $p < 0.001$ (extremely significant), $p < 0.01$ (very significant), and $p < 0.05$ (significant), respectively .

Index Terms: mass boundary detection, mammography, probability-based cost function

I. INTRODUCTION

In the United States, breast cancer accounts for one-third of all cancer diagnoses among women and it has the second highest mortality rate of all cancer deaths in women[1]. In several studies it has been shown that only 13% - 29% of suspicious masses were determined to be malignant[2-4], which indicates that there are high false positive rates for biopsied breast masses. A higher predictive rate is anticipated by combining the mammographer's interpretation and the computer analysis. Other studies have shown that 7.6% - 14% of the patients have mammograms that produce false negative diagnoses[5-6]. Alternatively, a Computer Assisted Diagnosis (CAD_x) system can serve as a clinical tool for the radiologist and consequently lower the rate of missed breast cancer.

Generally, CAD_x systems consist of three major stages, namely, segmentation, feature calculation, and classification. Segmentation is arguably one of the most important aspects of CAD_x – particularly for masses - because a strong diagnostic predictor for masses is shape. Specifically, many malignant masses have ill-defined, and/or spiculated borders and many benign masses have well-defined, rounded borders. Furthermore, breast masses can have unclear borders and are sometimes obscured by glandular tissue in mammograms. During the search for suspicious areas it is possible that masses of this type are overlooked by radiologists. When a specific area is deemed to be suspicious, the radiologist analyzes the overall mass, including its shape and margin characteristics. The margin of a mass is defined as the interface between the mass and surrounding tissue, and is regarded by some as one of the most important factors in determining its significance[7]. Specifically, a spiculated mass consists of a central mass body surrounded by fibrous extensions, hence the resulting stellate shape. In this context, "extension" refers to those portions of the mass containing ill-defined borders, spiculations, fibrous borders, and projections. Although the diameters of these cancers are measured across

the central portion of the mass, microscopic analysis of the extensions also reveals associated cancer cells, i.e., the extended projections may contain active mass growth[7,8]. In addition, the features of the extended projections and ill-defined borders are highly useful for identifying masses. Hence, proper segmentation - to include the body and periphery - is extremely important and is essential for the computer to analyze, and in turn, determine the malignancy of the mass in mammographic CAD_x systems.

Te Brake and Karssemeijer[9] implemented a discrete dynamic contour model, a method similar to snakes, which begins as a set of vertices connected by edges (initial contour) and grows subject to internal and external forces. Li[10] developed a method that employs k-means classification to classify pixels as belonging to the region of interest (ROI) or background. Petrick et al.[11] developed the Density Weighted Contrast Enhancement (DWCE) method, in which series of filters are applied to the image in an attempt to extract masses. Pohlman et al.[12] developed an adaptive region growing method whose similarity criterion is determined from calculations made in 5×5 windows surrounding the pixel of interest. Mendez et al.[13] developed a method, which combined bilateral image subtraction and region growing to segment masses.

Several studies have also used probability-based analysis to segment digitized mammograms. Li et al.[14] developed a segmentation method that first models the histogram of mammograms using a finite generalized Gaussian mixture (FGGM) and then uses a contextual Bayesian relaxation labeling (CBRL) technique to find suspected masses. Furthermore, this method uses the Expectation-Maximization (EM) technique in developing the FGGM model. Comer et al.[15] utilized an EM technique to segment digitized mammograms into homogeneous texture regions by assigning each pixel was to one of a set of classes such that the number incorrectly classified pixels was minimized. Kupinski and Giger[16] developed a method, which combines region growing with probability analysis to determine final segmentation. In their

method, the probability-based function is formed from a specific composed probability density function, determined by a set of image contours produced by the region growing method. This method is a highly effective one and it was implemented by Te Brake and Karssemeijer in their work[9] that compared the results of a model of the discrete dynamic contour model with those of the probability-based method. For this reason we chose to investigate its use as a possible starting point from which a second method could be developed. Consequently for our implementation of this work we discovered an important result, i.e., the steepest changes of a cost function composed from two probability density functions of the regions. It appears that in many cases this result produces contour choices that encapsulate important borders such as mass spiculations and ill-defined borders.

Several CAD_x classification techniques have been developed. They are described here to underscore the importance of accurate segmentation in CAD_x studies. Lo et al.[17] has developed an effective analysis method using the circular path neural network technique that was specifically designed to classify the segmented objects and can certainly be extended for the applications related to mass classification. Polakowski et al.[18] used a multilayer perceptron (MLP) neural network to distinguish malignant and benign masses. Both Sahiner et al.[19] and Rangayyan et al.[20] used linear discriminant analysis to distinguish benign masses from malignant masses. While many CAD_x systems have been developed, the development of fully-automated image segmentation algorithms for breast masses has proven to be a daunting task.

II. METHODS

A. Segmentation method – Maximum change of cost function as a continuation of probability-based function analysis

As a point of clarification, in this work we refer to the function used to find optimal region growing contours in Kupinski and Giger's study[16] as the probability-based function and we refer our function as the cost function. The two functions are similar; however they differ in terms of the images used in their formation. As an initial segmentation step, region growing is used to aggregate the area of interest[12, 13, 21], where grayscale intensity is the similarity criterion. This phase of the algorithm starts with seed point whose intensity is high, and nearby pixels with values greater than or equal to this value are included in the region of interest. As the intensity threshold decreases, the region increases in size, therefore there is an inverse relationship between intensity value and contour size. In many cases the region growing method is extremely effective in producing contours that are excellent delineations of mammographic masses. However, the computer is not able to choose the contour that is most highly correlated with the experts' delineations, specifically, those masses that contain ill-defined margins or margins that extend into surrounding fibroglandular tissue. Furthermore, the task of asking a radiologist to visually choose the best contour would be both time intensive and extremely subjective from one radiologist to another.

The segmentation technique described in this work attempts to solve and automate this process by adding a two-dimensional (2D) shadow and probability-based components to the segmentation algorithm. Furthermore, we have devised a steepest descent change analysis method that chooses the best contour that delineates the mass body contour as well as its extended borders, i.e., extensions into spiculations and areas in which the borders are ill-defined or obscured. It has been discovered that the probability-based function is capable of extracting the central portion of the mass density as demonstrated by the previous investigators[16], and in this work the method has been advanced further such that it can include the extensions of the masses. The enhanced method can produce contours, which closely match expert radiologist

traces. Specifically, it has been observed that this technique can select the contour that accurately represents the mass body contour for a given set of parameters. However, further analysis of the cost function composed from the probability density functions inside and outside of a given contour revealed that the computer could choose a set of three segmentation contour choices from the entire set of contour choices, and later make a final decision from these three choices.

1. Region growing and pre-processing

Initially, a 512×512 pixel area surrounding the mass is cropped. The region growing technique[12, 13, 21] to aggregate the region of interest was employed, where the similarity criterion for our region growing algorithm is grayscale intensity. To start the growth of first region, a seed point was placed at the center of the 512×512 ROI. The region growing process continues by decreasing the intensity value until we have grown a sufficiently large set of contours.

Next, the image is multiplied by a 2D trapezoidal membership function with rounded corners whose upper base measures 40 pixels and lower base measures 250 pixels (1 pixel = 50 microns). This function was chosen because it is a good model of the mammographic mass's intensity distribution. Since the ROI's have been cropped such that the mass's center was located at the center of the 512 pixel × 512 pixel area, shadow multiplication emphasizes pixel values at the center of the ROI and suppresses background pixels. The image to which the shadow has been applied is henceforth referred to as the "processed" image. The original image and its processed version were used to compute the highest possibility of its boundaries. The computation method is comprised of two components for a given boundary: (1) formulation of the composed probability as a cost function and (2) evaluation of the cost function.

The contours were grown using the original image as opposed to the processed image and this accounts for a major difference between the current implementation and that implemented by the previous investigators[16]. By using contours generated from the original image a cost function composed from the probability density functions inside and outside of the contours was produced. In many situations, the greatest changes in contour shape and size occur at sudden decreases within the function. In analyzing these steep changes it was observed that the intensity values corresponding to the steep changes typically produced contours that encapsulated both the mass body as well as its spiculated projections or ill-defined margins. This phenomenon would be suppressed if the processed image was used to generate the contour. A more detailed discussion of steep changes within the cost function is forthcoming in section II.A.2.3.

The processed image was mainly used to construct the cost function. A common technique used in mass segmentation studies is to pre-process the images using some type of filtering mechanism[11,16] in an effort to separate the mass from surrounding fibroglandular tissue. This method could be particularly beneficial to the region growing process because it would aid in preventing the regions from growing into surrounding tissue. Alternatively, the filtering process could impede our goal of attempting to encapsulate a mass's extended borders as well as borders that are ill-defined due to the filtering process's a tendency to create rounded edges on margins that are actually jagged, i.e., spiculated. This phenomenon could potentially defeat the goal of extracting mass borders. For these reasons, we have chosen to aggregate the contours using the original ROI rather its processed version.

2. Formulation of the composed probability as a cost function

In the context of this work, the composed probability is defined as the probability density functions of the pixels inside and outside a contour using a processed and non-processed version of an image. Specifically, for a contour (S_i), the composed probability (C_i) is calculated:

$$C_i \Big| S_i = \prod_{j=0}^{h} p\left(f_i(x,y) \middle| S_i\right) \times \prod_{j=0}^{h} p\left(m_i(x,y) \middle| S_i\right) \qquad (1)$$

The quantity $f_i(x,y)$ is the set of pixels, which lie inside the contour S_i (see Fig. 1a), and this area contained processed pixel values. The quantity $p(f_i(x,y)|S_i)$ is the probability density function of the pixels inside S_i ($f_i(x,y)$), where 'i' is the intensity threshold used to produce the contours given by the region growing step, and 'h' is the maximum intensity value. The quantity $m_i(x,y)$ is the set of pixels, which lie outside the contour S_i (see Fig. 1b), and this area contained non-processed pixels. The quantity $p(m_i(x,y)|S_i)$ is the probability density function of the pixels outside S_i, where 'i' is the intensity threshold used to produce the contours given by the region growing step, and 'h' is the maximum intensity value. For implementation purposes, the logarithm of the composed probability of the two regions, C_i was used.

$$Log\left(C_i \middle| S_i\right) = \log\left(\prod_{j=0}^{h} p\left(f_i(x,y) \middle| S_i\right)\right) + \log\left(\prod_{j=0}^{h} p\left(m_i(x,y) \middle| S_i\right)\right) \qquad (2)$$

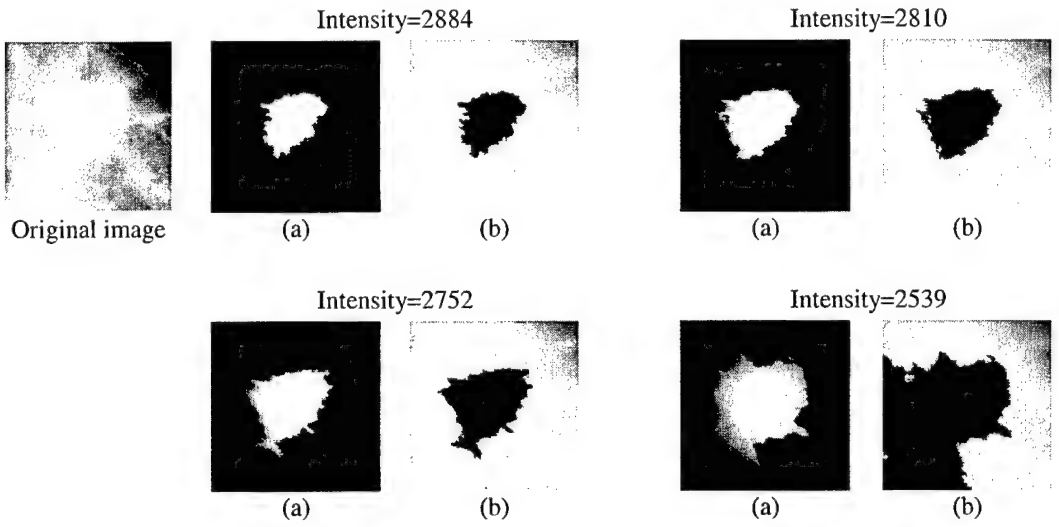


Fig. 1: Four grown contours used to construct the cost function : starts from high intensity thresholds and moves towards low intensity thresholds. Each contour separates the ROI into two parts: (a) Segmented image (based on processed image) used to compute density function $p(f_i(x,y)|S_i)$ and (b) Masked image (based on non-processed original image) used to compute density function $p(m_i(x,y)|S_i)$ for four intensity threshold values

3. The cost function based on the composed probability density functions

To select the contour that represents the fibrous portion of the mass, it is appropriate to examine the maximum value of the cost function:

$$\arg\max\left(Log\left(C_i|S_i\right), S_i, i=1,...n\right) \tag{3}$$

It has been assessed (also by other investigators[9,16]) that the intensity value corresponding to this maximum value is the optimal intensity needed to delineate the mass body contour. However, in the current implementation it was discovered that the intensity threshold corresponding to the maximum value confines the contour to the fibrous portion of the mass, i.e., the mass body. In the study many of these contours did not include the extended borders. It is therefore, hypothesized that the contour represents the mass's extended borders may well be determined by assessing the greatest changes of the cost function, i.e., locate the steepest value changes within the function:

$$\frac{d}{di}\left(Log\left(C_i|S_i\right), S_i, i=1,...n\right) \qquad (4)$$

Based on this assumption, the cost functions associated with masses were analyzed. The analysis reveals that the most likely boundaries of masses associated with expert radiologist traces are usually produced by the intensity value corresponding to the first or second steepest change of value immediately following the maximum value on the cost function (see Fig. 2a). The description of this discovery is given below followed by a validation study described in section II.B and results shown in section III. The overarching goal of the steep descent method is to determine the possibility that a certain contour is the best contour, which represents the mass and its extended borders.

3. The definition of steepest change

The term "steepest change" is rather subjective and in the context of this work can be defined as a location between two or more points in the cost function where the values experience a significant change. When the values are plotted as a function of intensity, these significant changes are often visible in the function. In some cases the cost function increases at a slow rate, therefore a potential steepest change location could be missed. The algorithm design compensates for this issue by calculating the difference between values in steps over several values and comparing the results to two threshold values. The difference equation is given by:

$$d(t) = f(z - wt) - f(z - w(t+1)), \qquad t = 0,\ldots,m \qquad (5)$$

where f is the cost function , z is the maximum intensity, w is the width of the interval over which the cost function differences are calculated (e.g. – for w=5 differences are calculated every 5 points), and m is the total number of points in the searchable area divided by w. Note that "wt" is associated with a specific contour "i" described earlier. If the value of $d(t)$ yields a value

greater than or equal to a given threshold, then the intensity corresponding to this location is determined to be a steepest change location. The threshold algorithm occurs as follows:

If $(d(t) \geq TV_1)$; t=0,…,m

Then choice 1 = intensity where that condition is satisfied.

If $(d(t) \geq TV_2)$; t=m,…,z

Then choice 2 = intensity where that condition is satisfied.

where TV_1 and TV_2 are pre-defined threshold values, m is the location in the function where the choice 1 condition is satisfied, and z is the location in the function where the choice 2 condition is satisfied. During the examination of the contour growth with respect to the cost function , the first steepest change (i.e., $d(t)_{MC1}$ as choice 1) is determined by TV_1 immediately after the location of the maximum cost function value (corresponding to mass body discussed earlier). The second the steepest change (i.e., $d(t)_{MC2}$ as choice 2) is determined by TV_2 after the first steepest change has been established.

As an example Fig. 1a is used to illustrate how the algorithm is carried out. In this figure, the maximum value on the cost function occurs for a grayscale intensity value of approximately 3330. The searching process begins from this maximum point and it is discovered that the first steepest change ($d(t)_{MC1}$ as choice 1) occurs for a grayscale intensity value approximately equal to 3200. From this point the continue the searching process continues and it is discovered that the second steepest change ($d(t)_{MC2}$ as choice 2) occurs for a grayscale intensity value approximately equal to 3175. In summary, intensity values of 3330, 3200, and 3175 can be used to grow 3 potential mass delineation candidates, and the large set of intensity choices has been narrowed to 3 choices. In many cases intensities, which produced the three contour choices gave the following results:

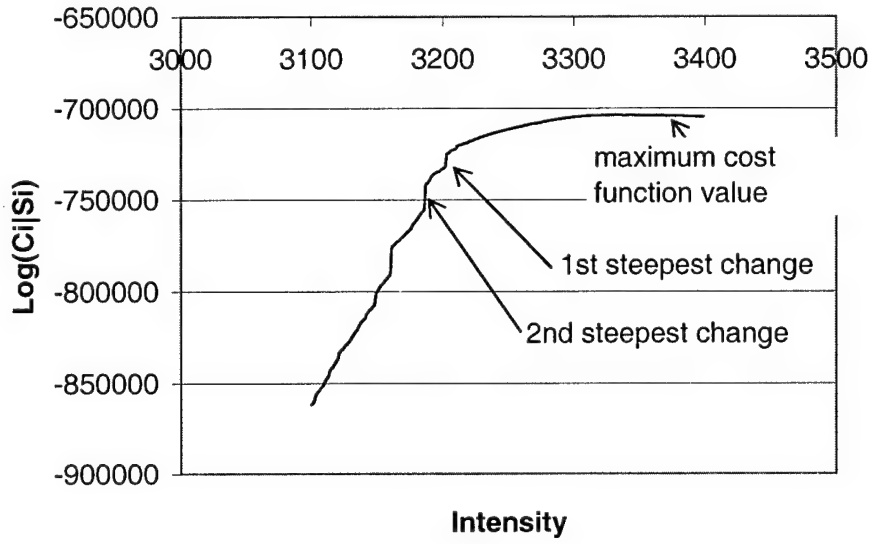
(1) Intensity corresponding to the maximum value on the cost function: The central body of the mass was encapsulated

(2) Intensity corresponding to the first steepest change on the cost function: The central body of the mass + some of its extended borders (i.e., projections and spiculations) was encapsulated

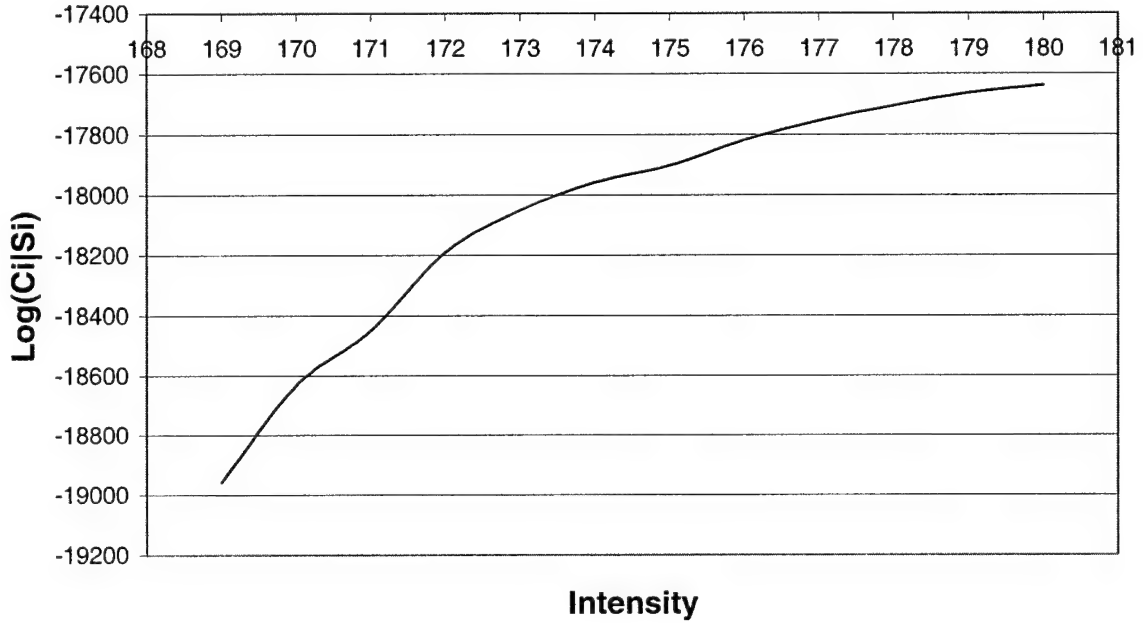
(3) Intensity corresponding to the second steepest change on the cost function: The central body of the mass + more extended borders + surrounding fibroglandular tissue encapsulated

The intensity corresponding to the first steepest change provides the best choice, and an examination of this observation is shown and discussed in sections III and IV of this work.

As stated previously the steep changes within the cost function would be suppressed if the processed image was used to generate the contour, therefore the function would be relatively smooth. This issue is evident in Fig. 2b, which shows a probability-based function produced by contours that were grown using a processed ROI.



(a)



(b)

Fig. 2: (a) Example of cost function with steepest change location indicators (b) Example of a probability-based function without an obvious steepest change location.

B. Validation method

In several segmentation studies the results were validated using the overlap statistic alone, however, it was necessary to analyze the performance of the steepest change algorithm on the basis of four statistics to verify that the algorithm is indeed capable of categorizing mass and background pixels correctly. This type of analysis provides helpful information regarding necessary changes for the algorithm's design and can possibly aid in its optimization.

The segmentation method was validated on the basis of overlap, accuracy, sensitivity, and specificity[22, 23]. These statistics are calculated as follows:

$$Overlap = \frac{E \cap P}{E \cup P} \tag{6}$$

$$Accuracy = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}} \tag{7}$$

$$Sensitivity = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{8}$$

$$Specificity = \frac{N_{TN}}{N_{TN} + N_{FP}} \tag{9}$$

where E is the drawing produced by the expert radiologist, P is the segmentation result, N_{TP} is the true positive fraction (part of the image correctly classified as mass), N_{TN} true negative fraction (part of the image correctly classified as surrounding tissue), N_{FP} is the false positive fraction (part of the image incorrectly classified as mass), and N_{FN} is the false negative fraction (part of the image incorrectly classified as surrounding tissue). This method requires a gold standard, or, contour to which the segmentation results can be compared. The gold standards for the experiments performed in this work were mass contours, which have been traced by expert radiologists.

The experiments produced contours for the intensity values resulting from three locations within the cost functions : (1) The intensity for which a value within the cost function is maximum (2) The intensity for which the cost function experiences its first steepest change and (3) The intensity for which the cost function experiences its second steepest change . It has been observed that the intensity for which the cost function experiences its first steepest change produces the contour trace that is most highly correlated with the gold standard traces, regarding overlap and accuracy. In cases for which better results occur at the second steepest change location, there is no significant difference between these results and the results calculated for the first steepest change location. Second, it has been observed that the results are more closely correlated with one expert than with the second expert. These hypotheses were tested using the one-way Analysis of Variance (ANOVA) test[24,25]. In this study, three significance levels (i.e., $p < 0.001$, $p < 0.01$, and $p < 0.05$) were used to categorize the ANOVA results as described in the next section.

III. EXPERIMENTS AND RESULTS

The following sections describe the database and experiments as well as provide results and ANOVA test results.

A. Database

For this study, a total of 124 masses were chosen from the University of South Florida's Digital Database for Screening Mammography (DDSM)[26]. The DDSM films were digitized at 43.5 or 50 μm's using either the Howtek or Lumisys digitizers, respectively. The DDSM cases have been ranked by expert radiologists on a scale from 1 to 5, where 1 represents the most subtle masses and 5 represents the most obvious masses. Table 1 lists the distribution of the masses

studied according to their subtlety ratings. The images were of varying contrasts and the masses were of varying sizes.

Table 1: Distribution of DDSM masses studied according to their subtlety ratings

Subtlety Category	Cancer	Benign
Number of masses with a rating = 1	5	3
Number of masses with a rating = 2	12	12
Number of masses with a rating = 3	18	17
Number of masses with a rating = 4	9	23
Number of masses with a rating = 5	15	10

The first set of expert traces was provided by an attending physician of the GUMC, and is hereafter referred to as the Expert A traces. The second set of expert traces was provided by the DDSM, and is hereafter referred to as the Expert B traces.

B. Experiments

As mentioned previously, the term "steepest change" is very subjective and therefore a set of thresholds needed to be set in an effort to define a particular location within the cost function as a "steepest change location". For this study the following thresholds were experimentally chosen: TV_1=1800, TV_2=1300, where TV_1= threshold for steepest change location 1 for the cost function, and TV_2 = threshold for steepest change location 2 for the cost function . A number of experiments were performed in an effort to prove that (1) the intensity for which the cost function experiences the first steepest change location produces the contour trace, which is most highly correlated with the gold standard traces with regard to overlap and accuracy. In cases for which the second steepest change location achieves better results, there are no significant differences between the values obtained from the first steepest change location and the second steepest change location. The experiments linked with these hypotheses comprise the studies for a single observer. We have also set out to prove that (2) our results are more closely correlated with one expert than with the second expert. The experiments linked with

this hypothesis comprise the studies between two observers. First segmentation results for two malignant cases are presented, followed by segmentation results for two benign cases. Second, the ANOVA results for a set of hypotheses are presented. The contours produced by the maximum value as well as the steepest change locations within the cost functions are labeled as follows:

(1) group 1: The intensity for which a value within the cost function is maximum

(2) group 2: The intensity for which the cost function experiences its first steepest change

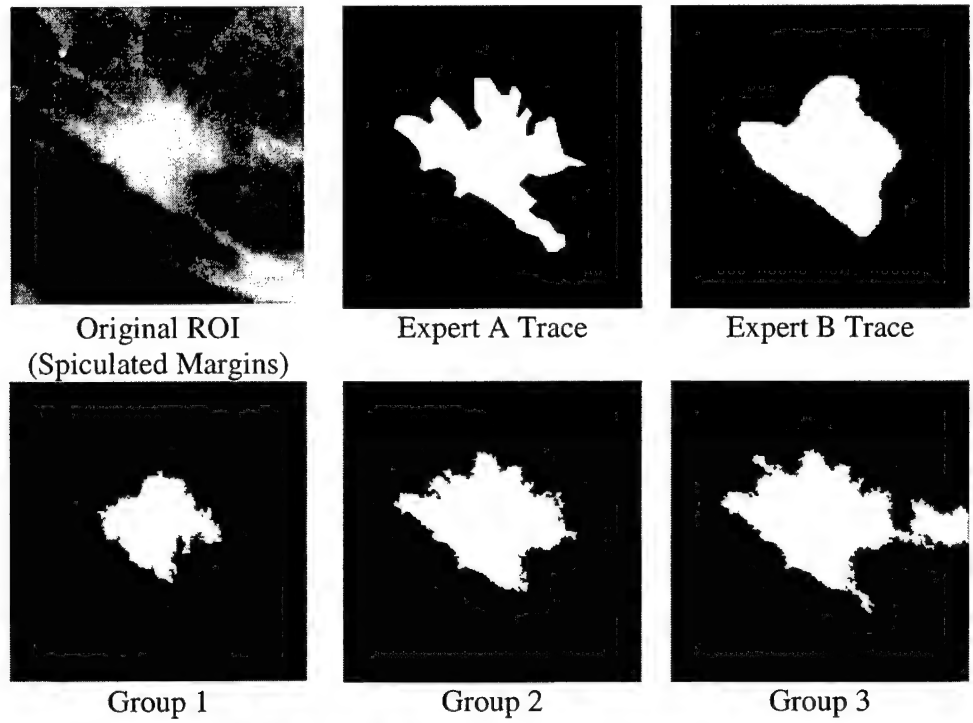
(3) group 3: The intensity for which the cost function experiences its second steepest change.

C. Results

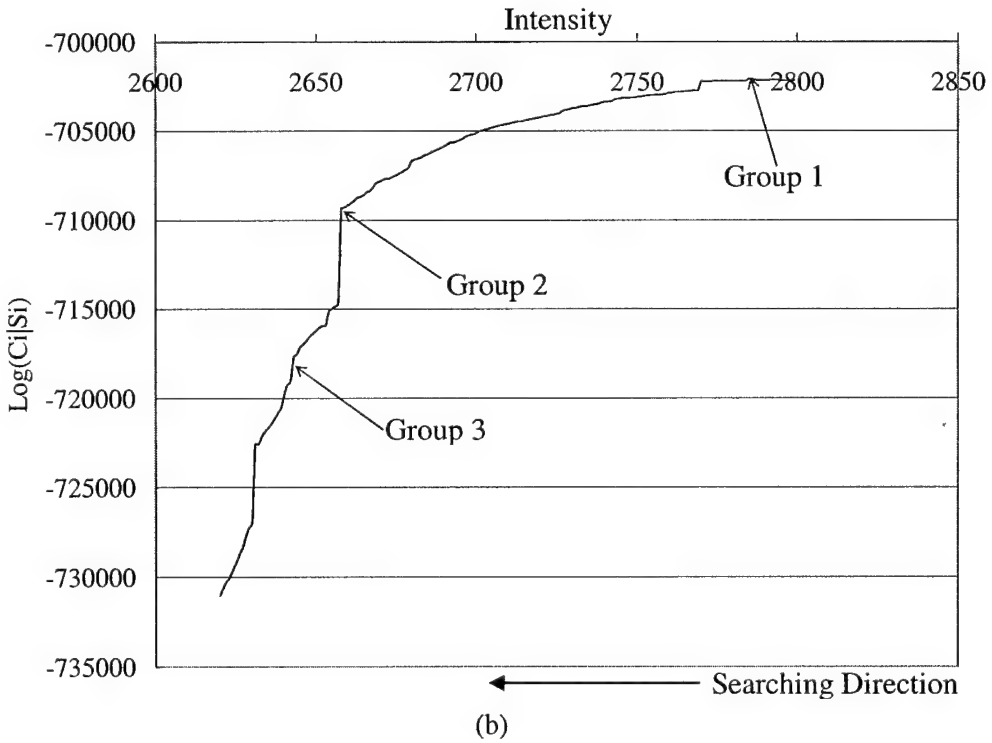
Figures 3-6 display the results for two malignant cases accompanied by their cost functions and results for two benign cases accompanied by their cost functions . The ANOVA results appear in a set of tables (sections 2-4), where each table lists the hypothesis tested along with p-values and their corresponding categorizations. The p-values are categorized in the following way: not significant (NS for $p > 0.05$), significant (S for $p < 0.05$), very significant (VS for $p<0.01$), and extremely significant (ES for $p < 0.001$). Each p-value table is followed by a second table, which contains the mean values of overlap, accuracy, sensitivity, and specificity for each group. Sections 2 and 3 are identical regarding the experiments, however, the pathologies of the masses are different (section 2 – malignant masses, section 3 – benign masses). Although the experiments are identical they have been separated for clarity purposes.

A larger set of segmentation results has been placed in an image gallery containing 7 malignant mass results (Fig. A1) and 7 benign mass results (Fig. A2). These figures are located in the Appendix.

1. Segmentation results

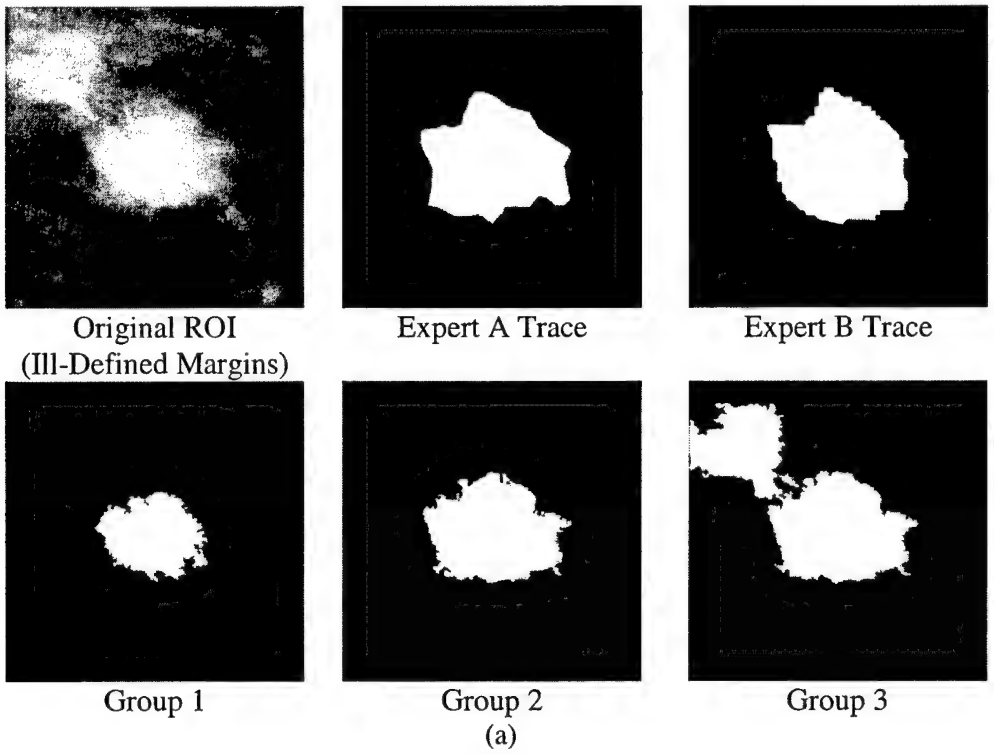


(a)



(b)

Fig. 3 – (a) Segmentation results for a malignant mass with spiculated margins (subtlety = 2)
(b) the corresponding cost function



(a)

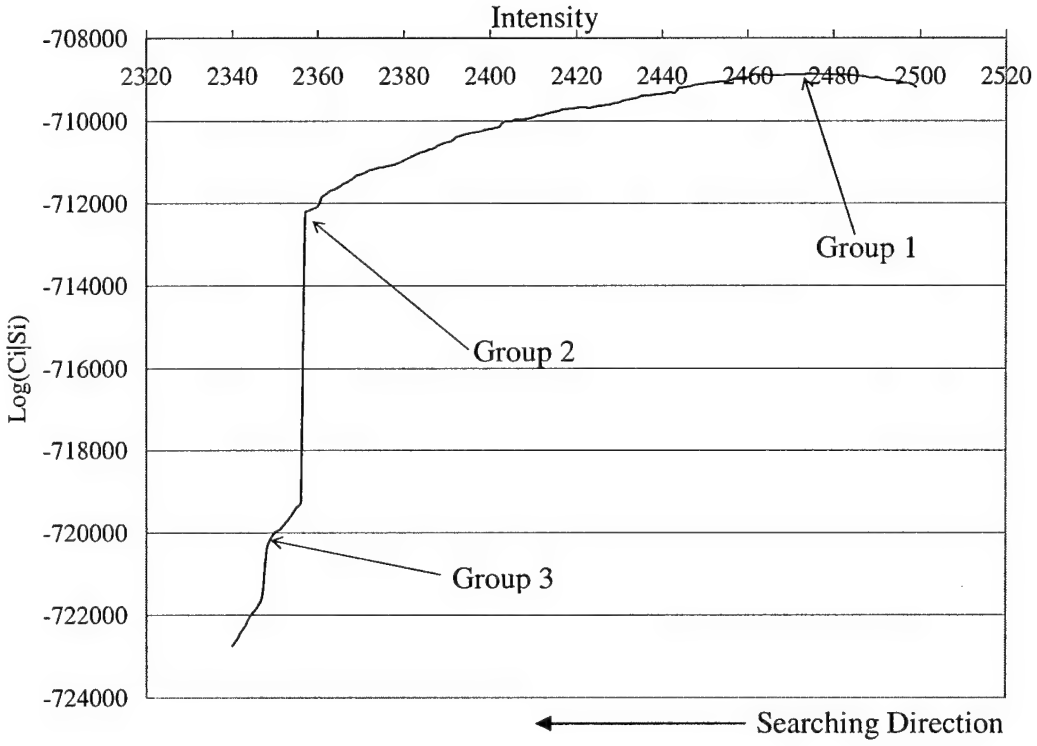
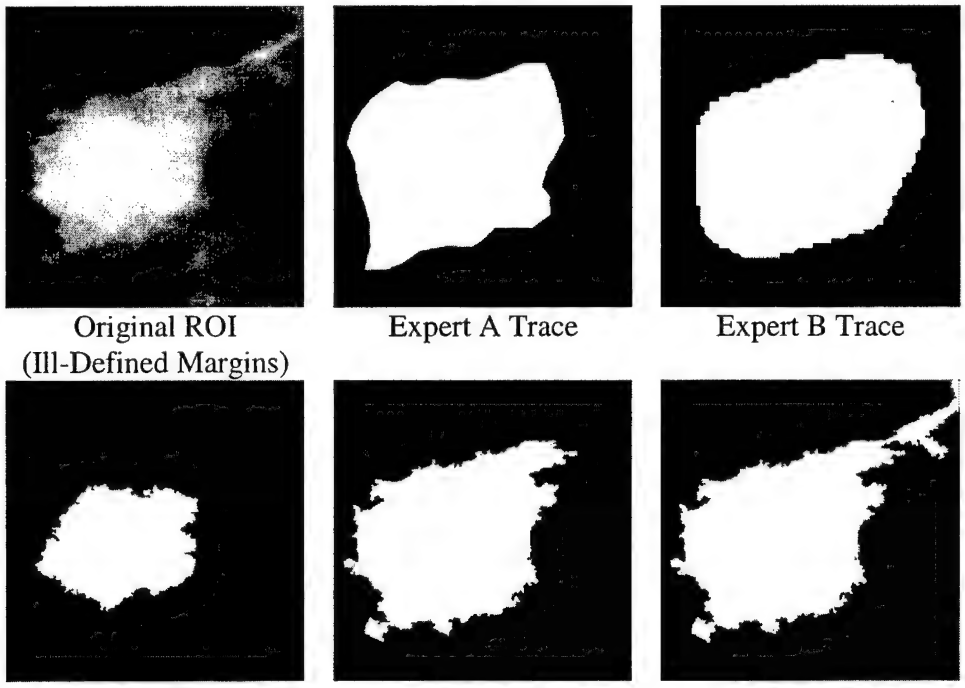


Fig. 4 – (a) Segmentation results for a malignant mass with ill-defined margins (subtlety = 3)
(b) the corresponding cost function



(a)

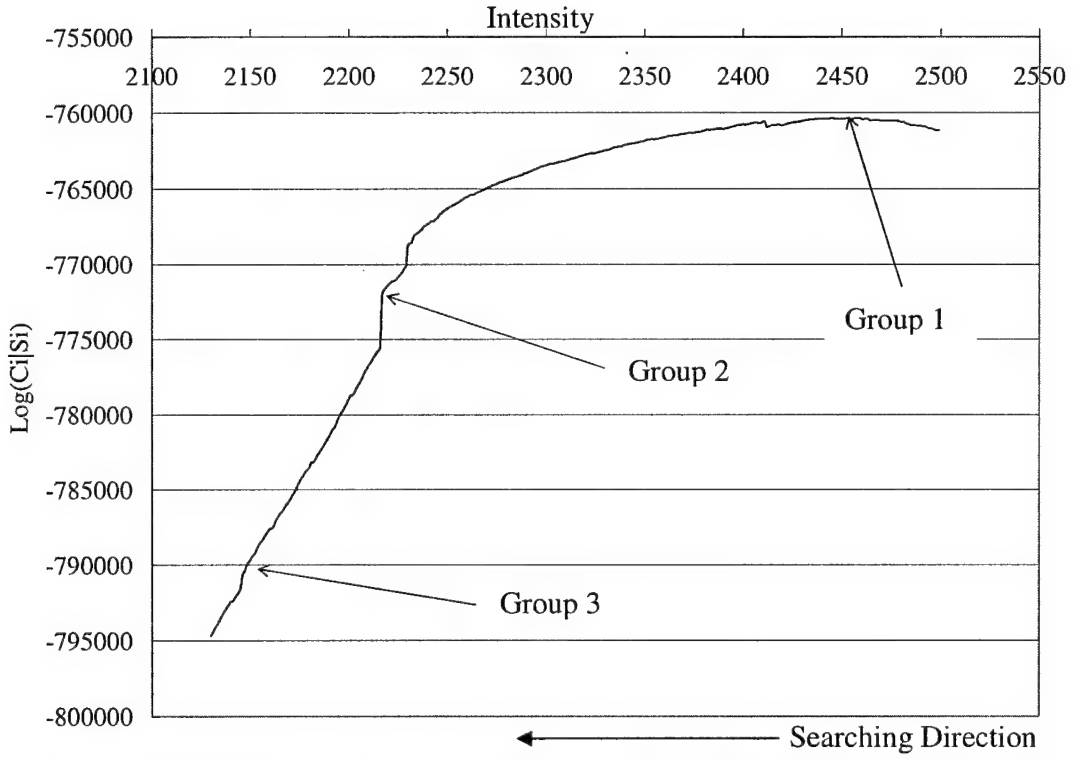
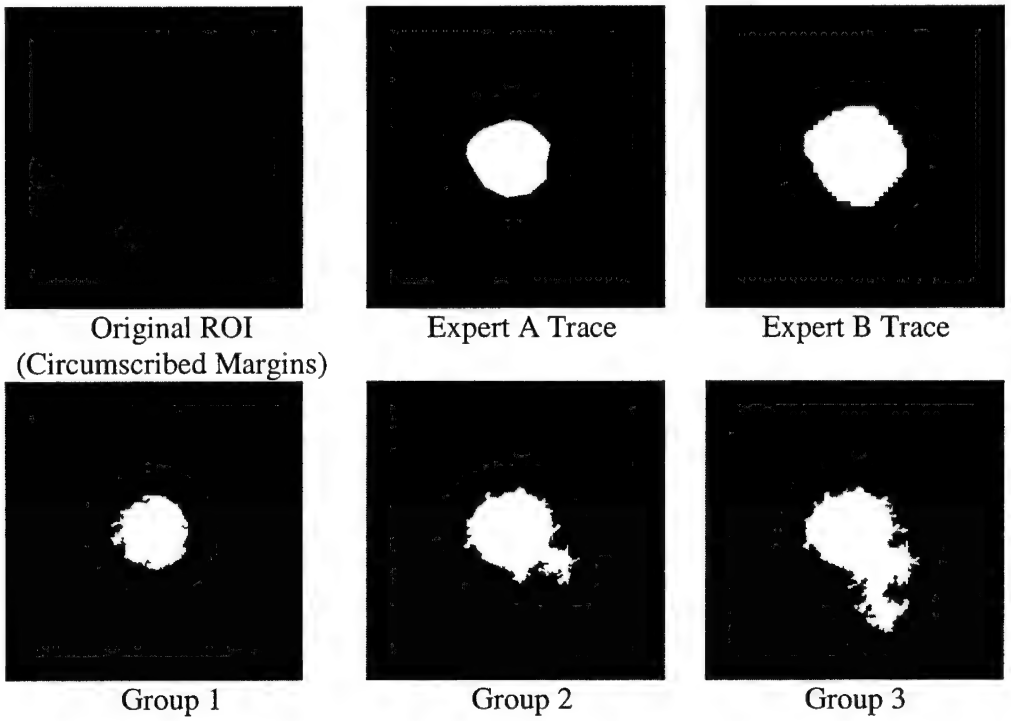


Fig. 5 – (a) Segmentation results for a benign mass with ill-defined margins (subtlety = 3)
(b) the corresponding cost function



(a)

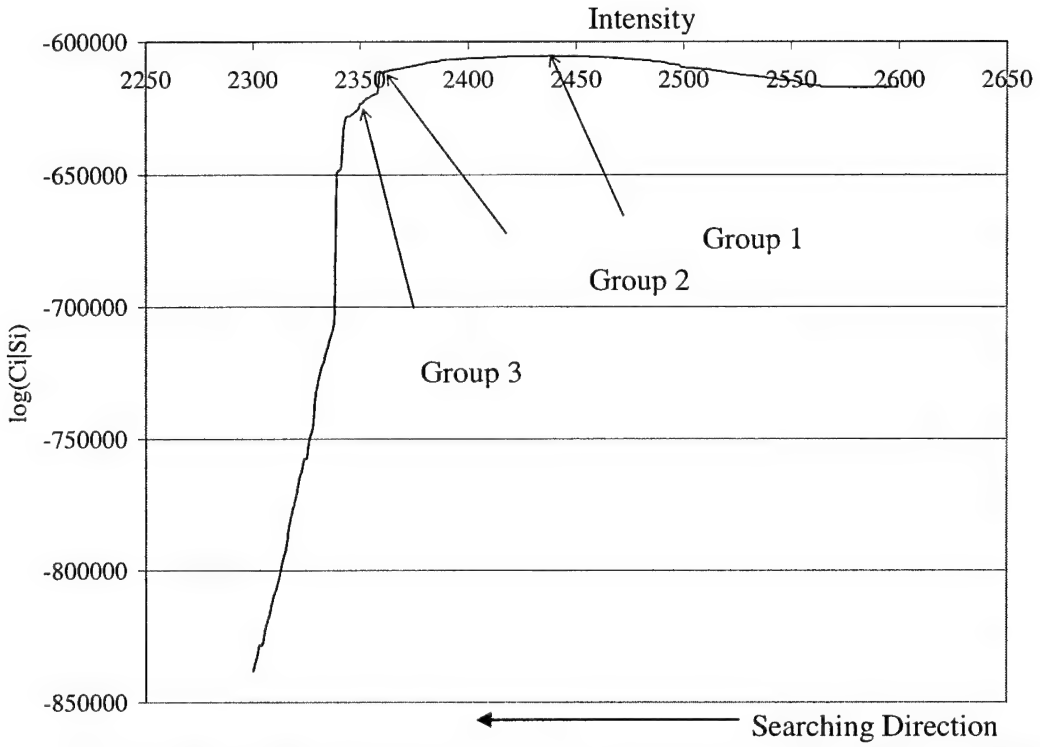


Fig. 6– (a) Segmentation results for a benign mass with circumscribed margins (subtlety = 4)
(b) the corresponding cost function

2. ANOVA test results for comparison of contour groups with single observer: malignant cases

Table 2: Single observer results (expert A gold standard, malignant masses)

ANOVA Test	P-value (group 1 vs. group 2)	P-value (group 2 vs. group 3)	P-value (group 1 vs. group 3)
Difference between groups (overlap)	1.78×10^{-4} (ES)	2.91×10^{-2} (S)	NS
Difference between groups (accuracy)	NS	3.14×10^{-2} (S)	NS
Difference between groups (sensitivity)	1.88×10^{-9} (ES)	NS	1.85×10^{-13} (ES)
Difference between groups (specificity)	5.12×10^{-4} (ES)	2.40×10^{-3}(VS)	2.71×10^{-9} (ES)

Table 3: Mean values for overlap, accuracy, sensitivity, and specificity (expert A gold standard, malignant masses)

Measurement	Mean Value (group 1)	Mean Value (group 2)	Mean Value (group 3)
Overlap	0.47	0.60	0.53
Accuracy	0.88	0.90	0.87
Sensitivity	0.49	0.75	0.81
Specificity	0.99	0.94	0.88

Table 4: Single observer results (expert B gold standard, malignant masses)

ANOVA Test	P-value (group 1 vs. group 2)	P-value (group 2 vs. group 3)	P-value (group 1 vs. group 3)
Difference between groups (overlap)	3.96×10^{-6} (ES)	NS	1.58×10^{-4}
Difference between groups (accuracy)	NS	NS	NS
Difference between groups (sensitivity)	4.88×10^{-8} (ES)	4.31×10^{-2} (S)	4.25×10^{-12} (ES)
Difference between groups (specificity)	2.70×10^{-4} (ES)	4.36×10^{-4} (ES)	1.44×10^{-7} (ES)

Table 5: Mean values for overlap, accuracy, sensitivity, and specificity (expert B gold standard, malignant masses)

Measurement	Mean Value (group 1)	Mean Value (group 2)	Mean Value (group 3)
Overlap	0.38	0.54	0.51
Accuracy	0.83	0.86	0.84
Sensitivity	0.38	0.56	0.60
Specificity	1.00	0.98	0.94

3. ANOVA test results for comparison of contour groups with single observer: benign cases

Table 6: Single observer results (expert A gold standard, benign masses)

ANOVA Test	P-value (group 1 vs. group 2)	P-value (group 2 vs. group 3)	P-value (group 1 vs. group 3)
Difference between groups (overlap)	3.19×10^{-4} (ES)	8.38×10^{-4} (ES)	NS
Difference between groups (accuracy)	NS	4.73×10^{-3} (VS)	2.51×10^{-3} (VS)
Difference between groups (sensitivity)	1.14×10^{-9} (ES)	1.89×10^{-2} (S)	7.51×10^{-17}(ES)
Difference between groups (specificity)	8.93×10^{-3} (VS)	1.24×10^{-3} (VS)	3.32×10^{-10} (ES)

Table 7: Mean values for overlap, accuracy, sensitivity, and specificity (expert A gold standard, benign masses)

Measurement	Mean Value (group 1)	Mean Value (group 2)	Mean Value (group 3)
Overlap	0.46	0.58	0.45
Accuracy	0.90	0.91	0.85
Sensitivity	0.49	0.73	0.82
Specificity	0.99	0.94	0.86

Table 8: Single observer results (expert B gold standard, benign masses)

ANOVA Test	P-value (group 1 vs. group 2)	P-value (group 2 vs. group 3)	P-value (group 1 vs. group 3)
Difference between groups (overlap)	8.82×10^{-5} (ES)	NS	1.62×10^{-2} (S)
Difference between groups (accuracy)	NS	2.62×10^{-2} (S)	2.48×10^{-2} (S)
Difference between groups (sensitivity)	1.61×10^{-7} (ES)	NS	3.14×10^{-12} (ES)
Difference between groups (specificity)	1.18×10^{-2} (S)	1.27×10^{-2} (S)	1.25×10^{-7} (ES)

Table 9: Mean values for overlap, accuracy, sensitivity, and specificity (expert B gold standard, benign masses)

Measurement	Mean Value (group 1)	Mean Value (group 2)	Mean Value (group 3)
Overlap	0.36	0.51	0.44
Accuracy	0.88	0.89	0.83
Sensitivity	0.36	0.61	0.69
Specificity	0.99	0.94	0.86

4. ANOVA test results for comparison of contour groups between two observers

Table 10: Two observer results: expert A vs. expert B, malignant masses

ANOVA Test	P-value (group 1 vs. group 2)	P-value (group 2 vs. group 3)	P-value (group 1 vs. group 3)
Expert A vs. Expert B (overlap)	3.12×10^{-3} (VS)	3.32×10^{-2} (S)	NS
Expert A vs. Expert B (accuracy)	1.20×10^{-2} (S)	4.46×10^{-2} (S)	NS
Expert A vs. Expert B (sensitivity)	9.43×10^{-4} (ES)	3.38×10^{-4} (ES)	3.67×10^{-4} (ES)
Expert A vs. Expert B (specificity)	NS	NS	NS

Table 11: Mean values for overlap, accuracy, sensitivity, and specificity (expert A vs. expert B, malignant masses)

Measurement	Mean Value, Expert A (group 1)	Mean Value, Expert B (group 1)	Mean Value, Expert A (group 2)	Mean Value, Expert B (group 2)	Mean Value, Expert A (group 3)	Mean Value, Expert B (group 3)
Overlap	0.49	0.38	0.62	0.55	0.55	0.51
Accuracy	0.89	0.83	0.91	0.87	0.87	0.84
Sensitivity	0.52	0.38	0.75	0.60	0.82	0.68
Specificity	0.99	1.00	0.95	0.97	0.89	0.91

Table 12: Two observer results: expert A vs. expert B, benign masses

ANOVA Test	P-value (group 1 vs. group 2)	P-value (group 2 vs. group 3)	P-value (group 1 vs. group 3)
Expert A vs. Expert B (overlap)	NS	NS	NS
Expert A vs. Expert B (accuracy)	NS	NS	NS
Expert A vs. Expert B (sensitivity)	3.56×10^{-2} (S)	4.90×10^{-2} (S)	2.03×10^{-2} (S)
Expert A vs. Expert B (specificity)	NS	NS	NS

Table 13: Mean values for overlap, accuracy, sensitivity, and specificity: expert A vs. expert B, benign masses

Measurement	Mean Value, Expert A (group 1)	Mean Value, Expert B (group 1)	Mean Value, Expert A (group 2)	Mean Value, Expert B (group 2)	Mean Value, Expert A (group 3)	Mean Value, Expert B (group 3)
Overlap	0.42	0.35	0.57	0.50	0.48	0.44
Accuracy	0.90	0.88	0.91	0.89	0.85	0.83
Sensitivity	0.44	0.36	0.71	0.61	0.79	0.69
Specificity	0.99	0.99	0.94	0.94	0.86	0.86

IV. DISCUSSION

A. Segmentation Results

From the ROI's shown in Figures 3 and 4 it is evident that the intensity produced by the maximum value is capable of accurately delineating the mass body contour, and in some cases this intensity corresponding to the maximum value produces a contour, which falls inside the mass body contour. This can be potentially problematic because low segmentation sensitivities can produce large errors during the feature calculation and diagnosis phases of CAD_x. Of the three available segmentation choices for each mass, it appears that the first steepest change location produces the contours with strongest correlation in comparison to both gold standards. These contours appear to cover both the mass body contour as well as the extended borders. In some instances the region grows into some areas that are not declared as mass areas by the gold standards – we call this flooding - and fails to grow into other areas that have been declared as mass areas. Finally, the second steepest change location produces contours that also cover both the mass body contour as well as the extended borders, and, the contours tend to also include surrounding fibroglandular tissue; hence, the flooding phenomenon is a common occurrence. In the cases shown, it is clear that steepest change location 1 produces the best contours, in comparison to the gold standards, however the ANOVA test results allow us to make such a claim. The following discussion is divided into five sections: single observer malignant results, single observer benign results, and two observer results (malignant and benign), algorithm performance, and an additional discussion on methods.

B. Malignant Cases with Single Observer

For both the Expert A and Expert B gold standards, Tables 2-5 show a statistically significant difference between groups 1 and 2 on the basis of overlap and sensitivity, where the

mean values of group 2 were higher than the mean values of group 1 for these statistics. These results are expected because as shown in the figures, the group 2 contours consistently covered more of the mass area (and correctly covered this mass area) as compared to the group 1 contours, according to both experts. There was a statistically significant difference in sensitivity between group 1 and group 3, where the mean of group 3 was higher than the mean of group 1. This is an expected result because out of all the groups, group 3 contours consistently cover the most mass area. For the Expert B gold standard there was a statistically significant difference in overlap between group 1 and group 3, where the mean of group 3 was higher than the mean of group 1. This is an expected result because out of all the groups, group 3 contours correctly cover the most mass area.

C. Benign Cases with Single Observer

For the Expert A there were statistically significant differences between the group 2 and group 3 traces on the basis of overlap, accuracy, and sensitivity, where the group 2 mean values for overlap and accuracy were higher than those of group 3 (see Tables 6-9). This is an expected result because it is likely that many of the group 3 contours contained flooded areas, which will cause both of these values to be lower than contours without flooded areas. The overlap and sensitivity values for group 2 were significantly higher than those of group 1. This is an expected result because the group 2 contours not only covered more mass area and correctly cover this area. Finally, the group 3 accuracy and sensitivity values were significantly higher than those for group 1. This is an expected result because the group 3 contours not only cover more mass area but also correctly cover this area.

For the Expert B gold standard there were statistically significant differences between the group 2 and group 3 traces on the basis of accuracy and sensitivity, where the group 2 mean

values for overlap and accuracy were higher than those of group 3. This is an expected result because it is likely that many of the group 3 contours contained flooded areas, which will cause both of these values to be lower than contours without flooded areas. There were statistically significant differences between group 1 and group 2 on the basis of overlap and sensitivity, where the mean values for group 2 were higher than the mean values for group 1. This is an expected result because the group 2 contours not only cover more mass area and correctly cover this area. There were statistically significant differences between group 3 and group 1 on the basis of overlap and sensitivity, where the mean values for group 3 were higher than those of group 1. This is an expected result because the group 3 contours not only covered more mass area and correctly covered this area.

In nearly all cases for the single observer studies, it was expected that the specificity values for group 1 would always be higher for group 1 than those for groups 2 and 3 because this contour always covered the smallest mass area, consequently its background was always highly correlated with the background areas dictated by the gold standards. Moreover, in some cases the group 2 and group 3 contours grew into areas that were not regarded as mass, but rather were regarded as background, therefore their specificity values had a lower correlation with the gold standard as compared to the group 1 contours.

D. Malignant and Benign Cases with Two Observers

For the two observer studies, comparisons were made between experts A and B on a group-by-group basis in an effort to prove that there were significant differences between the two radiologists on the basis of overlap, accuracy, sensitivity, and specificity (see Tables 10-13). For the malignant masses there were statistically significant differences between the two experts on the basis of overlap, accuracy, and sensitivity. There was a statistically significant

difference between the two experts for group 3 on the basis of sensitivity. For the benign masses, there were statistically significant differences between the two experts for all three groups on the basis of sensitivity. For all cases Expert A's values were consistently higher than those of Expert B. It is an expected result that there were statistically significant differences between the experts due to their differences in opinion. The fact that Expert A's mean values were higher than those for expert B, however does not warrant the conclusion that Expert A is a more reliable expert; however it does not warrant the conclusion that there is stronger agreement between the computer's results and Expert A's traces. Further, there were less statistically significant differences for the benign cases than for the malignant cases. This is an expected result because in general, benign masses have better defined borders so it was expected that the two experts would strongly agree.

E. Algorithm performance

It appears that the thresholds chosen produce first steepest change location intensities that generate contours that are closely correlated with the expert traces. In some instances the second steepest change location is extremely far from the first steepest change location, which implies that the function in question increases very slowly; and, many of the second steepest change location intensities produce contours with flooded areas. For the majority of the cases in which the second steepest change location contour achieves a higher sensitivity value, but not a significantly higher sensitivity value, we can still choose the first steepest change location contour because the difference between the two contours is likely to be negligible.

In analyzing the probability-based cost functions, we found that those functions with very steep changes are typically associated with masses that have well-defined borders while those functions that increase slowly are associated with masses that have ill-defined borders. This

phenomenon may make it necessary to develop an adaptive threshold process for the steepest change evaluation such that the functions are grouped into various categories (e.g. – smooth versus steep) because a threshold value that is optimal for the steep function may not be optimal for a smooth function.

F. Additional discussion on methods used

In this study it appears that the steepest descent method has the advantage of locating ill-defined margins as well as extensions such as malignant spiculations and projections for mammographic masses. If the human eye is solely used, it can be difficult to separate the mass from surrounding fibroglandular tissue. Therefore, it is believed that this method has the potential to complement the process of reading mammographic films. One of the downfalls of the method is its dependence upon the assumption that masses are generally light in color. This assumption impedes the region growing process because masses that contain darker areas and are surrounded on one or more sides by bright tissue can cause contours to flood into areas that are not actual mass tissue. Typically, this situation occurs for the mass located on the border of the breast region on a mammogram.

All of the segmentation methods surveyed in the introduction of this paper are excellent solutions for the problems the authors set out to solve, however, in some cases it is difficult to make comparisons between different methods without the availability of a set of several visual results. In several studies, the focus was either to detect masses or to distinguish malignant from benign masses. So the validation process did not take the form of a comparison with expert radiologist manual traces, but rather features were calculated on the potential mass candidates and they were later classified as being mass tissue or normal tissue[10,11,12,13]. The purpose of Li's study[14] was to distinguish normal and abnormal tissue so the authors did not

provide any statistics such as overlap or accuracy. Nevertheless, the study contains a figure of 60 masses that contain both computer and radiologist annotations to give the reader an idea of the computer algorithm's performance. Te Brake and Karssemeijer's study[9] used the overlap statistic to test the efficacy of their method and they indicated that the central mass area of the mass was delineated by the radiologist and their computer results were compared to these annotations. Kupinski and Giger's study[16] also used the overlap statistic to test the efficacy of their method and set a threshold for which the mass was considered to be successfully segmented. For example, masses whose overlap values are greater than 0.7 imply that there was successful segmentation.

The technical method presented herein shows that the results obtained from the maximization of the composed probability density function (i.e., the cost function) are equivalent to those obtained from previous methods presented by previous investigator. However, the steepest change of the composed probability density function is most close to the radiologist determination.

V. CONCLUSION

We have shown that our fully automatic boundary detection method for malignant and benign masses can effectively delineate these masses using intensities, which correspond to the first steepest change location within their cost functions . Additionally, it appears that the method is more highly correlated with one set of expert traces than with a second set of expert traces, regarding the accuracy and overlap statistics. This result shows that inter-observer variability can be an important factor in segmentation algorithm design, and it has motivated us to seek the opinions of more expert radiologists to test the robustness of our algorithm. The second steepest change location intensity will always yield contours with higher sensitivity

values, however, it behooves us to choose the first steepest change location intensity because it avoids the risk of choosing contours that contain substantial flooding. In future work, a worthwhile study would be to run the experiments for different threshold values in an effort to discover the possibility of deriving an optimal threshold procedure. We believe that such a procedure would improve the method of choosing optimal contours.

ACKNOWLEDGEMENTS

This work was supported by US Army Grant No.'s DAMD17-03-1-0314, DAMD17-01-1-0267, DAMD 17-00-1-0291 and DAAG55-98-1-0187. The authors would also like to thank the Referees for their constructive comments and recommendations.

REFERENCES

1. J.V. Lacey Jr, S.S. Devesa, L.A. Brinton, "Recent trends in breast cancer incidence and mortality", Environmental and Molecular Mutagenesis, **39**, no. 2-3, 82-88 (2002).

2. J.E. Meyer, D.B. Kopans, P.C. Stomper, K.K. Lindfors, "Occult Breast Abnormalities: Percutaneous Preoperative Needle Localization", Radiology, **150**, no. 2, 335-337 (1984).

3. A.L. Rosenberg, G.F. Schwartz, S.A. Feig, A.S. Patchefsky, "Clinically occult breast lesions: localization and significance", Radiology, **162**, 167-170 (1987).

4. B.C.Yankaskas, M.H. Knelson, M.L. Abernethy, J.T. Cuttino, R.L. Clark, "Needle localization biopsy of occult lesions of the breast", Investigative Radiology, **23**, 729-733 (1988).

5. J.A. Harvey, L.L. Fajardo, C.A. Innis, "Previous mammograms in patients with impalpable breast carcinoma: retrospective vs. blinded interpretation", American Journal of Roentgenology, **161**, 1167-1172 (1993).

6. J.E. Martin, M. Moskowitz, J.R. Milbrath, "Breast cancer missed by mammography", American Journal of Roentgenology, **132**, 737-739, (1979).

7. J.R. Harris, M.E. Lippman, M. Morrow, S. Hellman, "Diseases of the breast", Lippincott-Raven Publishers, Philadelphia, PA, 80-81 (1996).

8. J.E. Martin, "Atlas of mammography: histologic and mammographic correlations (second edition)", Williams and Wilkins, Baltimore, MD, p. 87 (1988).

9. G.M. te Brake, N. Karssemeijer, "Segmentation of suspicious densities in digital mammograms", Medical Physics, **28**, 259-266 (2001).

10. L. Li, Y. Zheng, L. Zhang, R. Clark, "False-positive reduction in CAD mass detection using a competitive classification strategy", Medical Physics, **28**, 250-258 (2001).

11. N. Petrick, H.-P. Chan, B. Sahiner, D. Wei, "An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection", IEEE Trans. on Med. Imag., **15**, 59-67 (1996).

12. S. Pohlman, K.A. Powell, N.A. Obuchowski, W.A. Chilcote, S. Grundfest-Broniatowski, "Quantitative classification of breast tumors in digitized mammograms", Medical Physics, **23**, no. 8, 1336-1345, (1996).

13. A.J. Méndez, P.G. Tahoces, M.J. Lado, M. Souto, J.J. Vidal, "Computer-aided diagnosis: Automatic detection of malignant masses in digitized mammograms", Medical Physics, **25**, 957-964 (1998).

14. H. Li, Y. Wang, K.J. R. Liu, S.-C. B. Lo, M.T. Freedman, "Computerized radiographic mass detection - part I: lesion site selection by morphological enhancement and contextual segmentation", IEEE Trans. on Med. Imag., **20**, 289-301 (2001).

15. M.L. Comer, S. Liu, E.J. Delp, "Statistical segmentation of mammograms", digital mammography '96: proceedings of the 3rd international workshop on digital mammography, Chicago, IL, 475-478 (9-12 June 1996).

16. M.A. Kupinski, M.L. Giger, "Automated seeded lesion segmentation on digital mammograms", IEEE Trans. on Med. Imag., **17**, no. 4, 510-517 (1998).

17. S.-C. B. Lo, H. Li, Y. Wang., L. Kinnard, and M.T. Freedman, "A multiple circular path convolution neural network system for detection of mammographic masses", IEEE Transactions on Medical Imaging, **21**, 150-158 (2002).

18. W.E. Polakowski, D.A. Cournoyer, S.K. Rogers , M.P. DeSimio, D.W. Ruck, J.W. Hoffmeister, R.A. Raines, "Computer-Aided Breast Cancer Detection and Diagnosis of Masses Using Difference of Gaussians and Derivative-Based Feature Saliency", IEEE Transactions on Medical Imaging, **16**, pp. 811-819 (1997).

19. B. Sahiner, H.-P. Chan, N. Petrick, M.A. Helvie, L.M. Hadjiiski, "Improvement of mammographic mass characterization using spiculation measures and morphological features", Medical Physics, **28**, pp. 1455-1465 (2001).

20. R.M. Rangayyan, N.M. El-Faramawy, J.E. Leo Desautels, O.A. Alim, "Measures of Acutance and Shape for Classification of Breast Tumors", IEEE Transactions on Medical Imaging, **16**, pp. 799-810 (1997).

21. B. Sahiner, H.-P. Chan, D.Wei, N. Petrick, M.A. Helvie, D.D. Adler, M.M. Goodsit, "Image feature selection by a genetic algorithm: Application to classification of mass and normal breast tissue", Medical Physics, **23**, pp. 1671-1684 (1996).

22. J. Suckling, D.R. Dance, E. Moskovic, D.J. Lewis, S.G. Blacker, "Segmentation of mammograms using multiple linked self-organizing neural networks", Medical Physics, **22**, 145-152, (1995).

23. B. Van Ginneken, "Automatic segmentation of lung fields in chest radiographs", Medical Physics, **27**, 2445-2455, (2000).

24. D. Downing , J. Clark, "Statistics the Easy Way, 2nd ed.", Hauppauge, NY: Barron's Educational Series, 184-206 (1989).

25. W. Hopkins, (2002) *A New View of Statistics: P values and statistical significance.* [Online]. Available: www.sportsci.org/resource/stats/pvalues.html.

26. M. Heath, K.W. Bowyer, D. Kopans et al., "Current status of the digital database for screening mammography", Digital Mammography, Kluwer Academic Publishers, 457-460 (1998).

APPENDIX A – Gallery of Segmentation Results

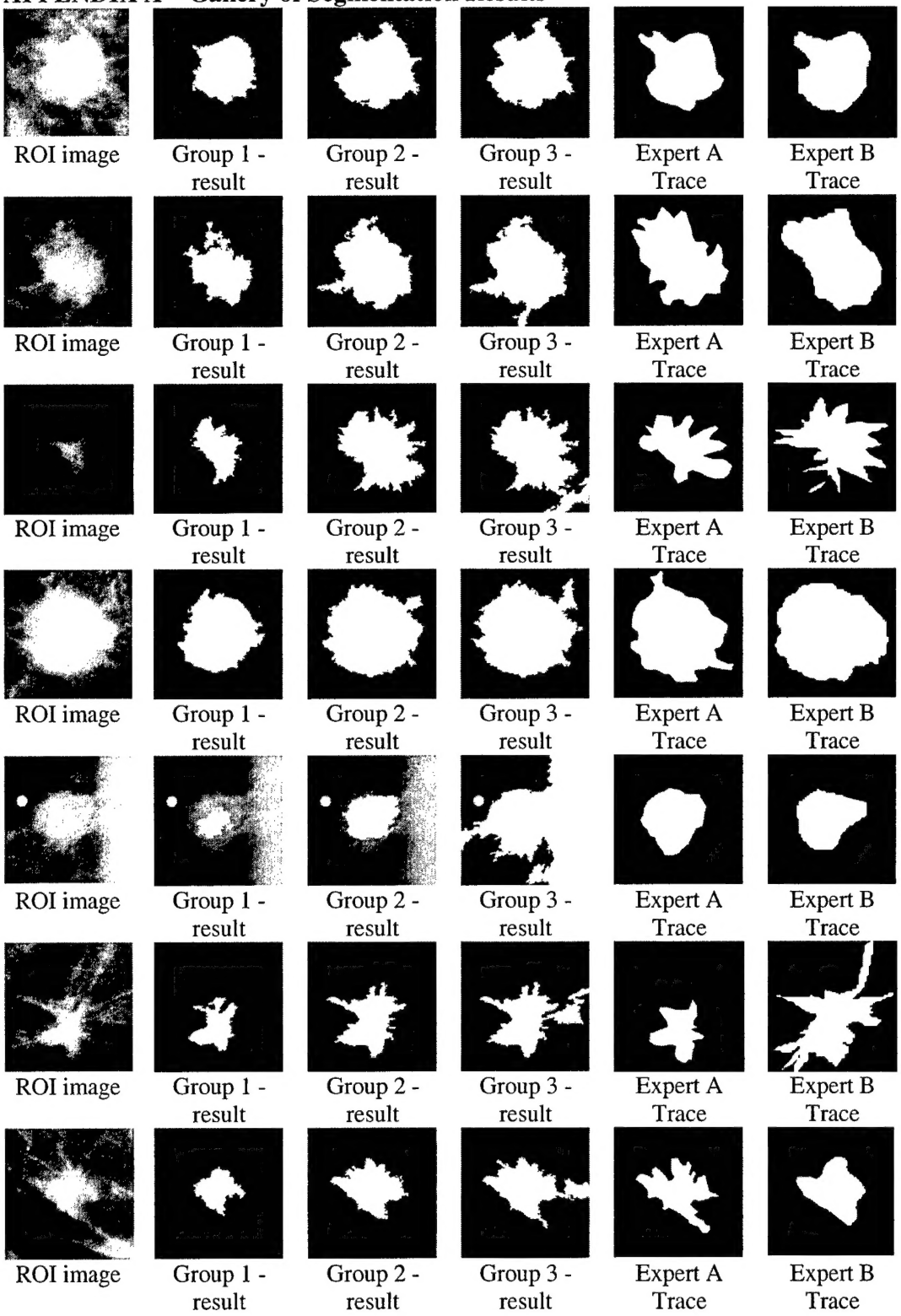


Fig. A1 – Segmentation results for a set of malignant masses

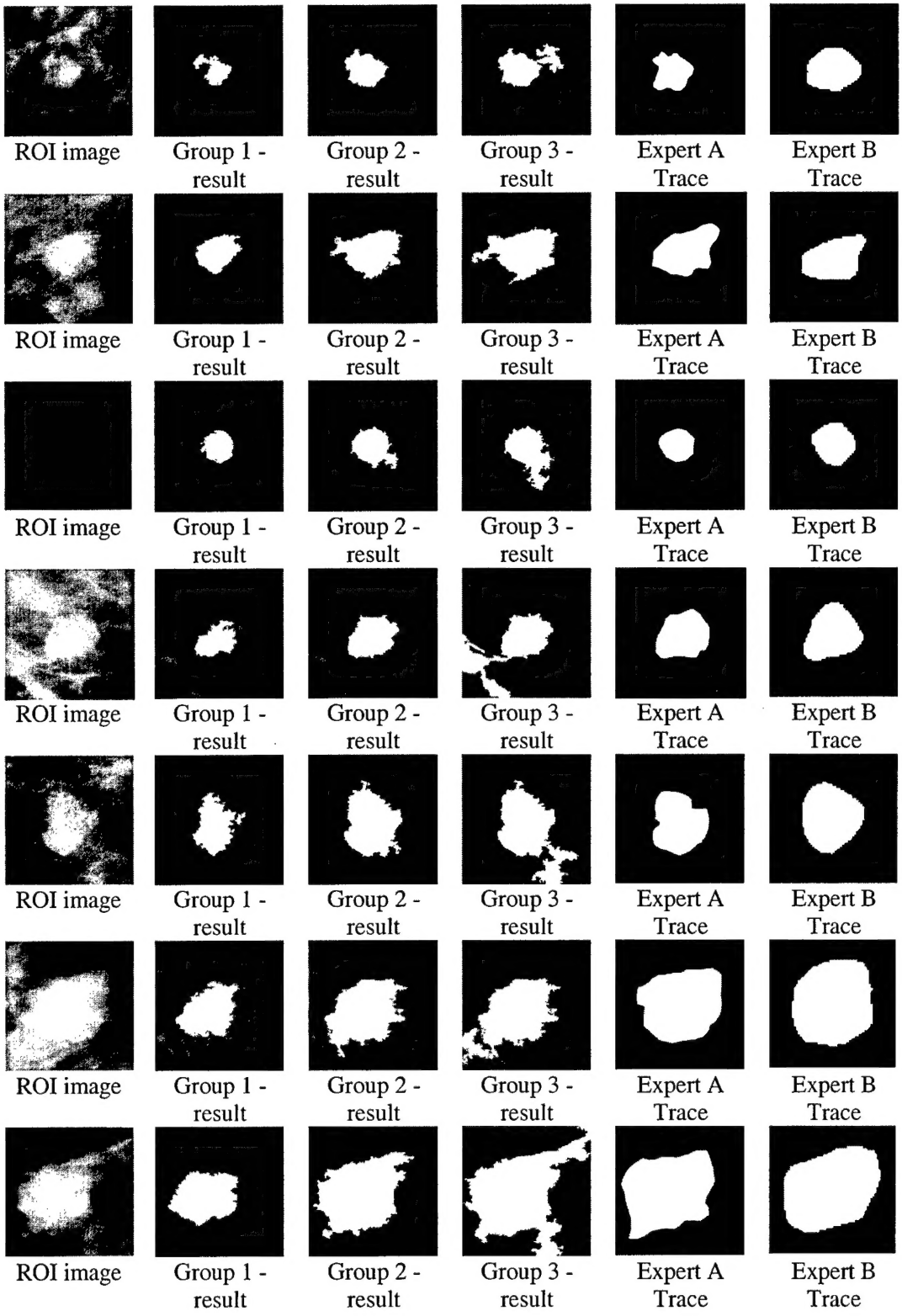


Fig. A2 – Segmentation results for a set of benign masses